



**United StatesTM
Census
Bureau**

DataFerrett Users Guide



Research & Methodology Directorate
DataWeb & Applications Staff

March 5, 2013

This page left blank intentionally.

DataFerrett Users Guide Overview

Chapter Number & Title	Topics Covered
Introduction	Getting started in DataFerrett Key features of DataFerrett Example of tasks in DataFerrett
Chapter 1: Selecting Variables & Adding Them to the DataBasket	How to find datasets Viewing available variables Using DataFerrett search capability Using “Browse/Select Variables” window to add variables to the DataBasket
Chapter 2: Selecting Geography	Using the geography codebook to select geographic areas
Chapter 3: Recodes and Other Things	How to create a recode variable Using ‘Subgroups repeat by..’ feature Recoding categorical value
Chapter 4: Make a Table	How to lay out a table How to create a table with selected data variables How to create derived columns Options for exporting the table out of DataFerrett
Chapter 5: Make a Graph	How to make a simple graph How to make time series graphs for datasets with “multiple instances”
Chapter 6: Make a Map	How to make a thematic map Options for changing the appearance of the map Printing and exporting the map
Chapter 7: Download an Extract from a Selected Dataset	Download formats available Limits on the size of a download Using the “Codebook” with the downloaded data Downloading in “batch” mode

This page left blank intentionally.

TABLE OF CONTENTS

INTRODUCTION	1
Getting Started in DataFerrett.....	2
CHAPTER 1: SELECTING VARIABLES & ADDING THEM TO THE DATABASKET	17
Adding Variables to the DataBasket	34
“Step 1” – Other Options.....	47
Menu Options	48
CHAPTER 2: BROWSE/SELECT GEOGRAPHY VARIABLES	51
Introduction to Ferrett Geography Codebook Interface.....	51
Geographic Selection Process Functionality.....	54
CHAPTER 3: RECODES & OTHER OPTIONS	69
What is a Recode?	69
Other Features of the Recode Window	76
Editing Labels for Recode Variables.....	79
Other Options in Step 2	84
CHAPTER 4: MAKE A TABLE.....	87
Ferrett Tabulation Tools	87
Make a Table: Basic Steps	89
The Ferrett Tabulation Screen.....	91
Ferrett Tabulation Menus	101
Creating New Columns Using Formulas.....	111
CHAPTER 5: MAKE A GRAPH	113
Ferrett Graph Window	113
Graph Types.....	126
CHAPTER 6: MAKE A MAP	139
Customizing a Map	145

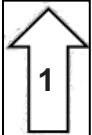
Changing Minimum and Maximum Values for the Ranges	146
Customizing Interval Breaks in the Legend	147
Changing Map Display and Legend to Quantile Intervals	148
Adding Layer Boundaries	148
CHAPTER 7: DOWNLOAD AN EXTRACT	151
Selecting File Formats	154
Getting Extracts	155

How to Use This Guide

The following symbols are used to assist you in understanding information and examples included in the Guide:



Indicates tips and helpful hints



Step number of example, shown on representation of user window

NOTE:

Alert related to specific example or explanation

Example table (shown below): Provides steps and information for each example

Example Table		
Steps	What to do	Why do it?
1, 2	Provides a brief description about what actions you need to take.	Describes what each action in the “What to do” section is accomplishing.

The complexity of an individual step may require more than a single action; in that case, sub-steps (“Step 1.1, 1.2, etc.”) are included followed by directions in the “What to do” column. The “Why do it?” column provides explanations/requirements for each step.

NOTE: The images used to illustrate user actions in Guide examples are based on the version of DataFerrett available at the time of publication. Updates to current datasets and DataFerrett functionality may cause slight inconsistencies between the Guide and the version of DataFerrett currently in use.

NOTE: Your feedback is important! Please send your suggestions for improvements to:
dsd_ferrett@census.gov

INTRODUCTION

DataFerrett is a web-based data analysis tool developed by the US Census Bureau designed to equip users with the ability to analyze large amounts of data and create customized reports to support decision making. Its users work for a myriad of organizations, such as agencies of the Federal Government, state and local governments, universities, private enterprise, non-profit organizations and divisions within the Bureau of the Census. DataFerrett draws upon the DataWeb, a network of public and private databases providing a vast amount of statistical information that is constantly updated and expanded.

The purpose of the DataFerrett Users Guide is to introduce the software to those who have no previous experience using it and to provide a quick reference manual for users who have used it but need information on specific features. The Guide will explain how become familiar with in DataFerrett, from logon to selecting datasets, variables and geographies and describe options for data analysis with specific examples that illustrate DataFerrett's many features. Also included is a cross-referenced index for finding information on particular features.

To maximize the benefits of this tool, it is important to understand its language. DataFerrett contains a repository of data from a variety of sources referred to as “**datasets**.” It enables a user to choose a dataset and/or a specific time period for a dataset called an “**instance**” and then select exact items called “**variables**” within the instance. In DataFerrett terminology, selecting the variables to work with is referred to as adding these variables to your “**DataBasket**.” DataFerrett offers a variety of actions that can be executed on the variables in the DataBasket to support data analysis.

NOTE: The datasets in DataFerrett come from many different sources and organizations that provide the data and the supporting documentation. As a result, some datasets are much more thoroughly documented than others.

There are two basic types of data that DataFerrett accesses:

1. **Microdata:** the data record that represents a survey response or an administrative record.
2. **Aggregated data:** a variable that contains a count of an estimate of a characteristic (e.g., the number of factories in a county or the number of people in the labor force).

At various stages in the DataFerrett process, there are differences between what can be done with a microdata dataset versus an aggregate dataset. These features are discussed in subsequent sections of the Guide.

Getting Started in DataFerrett

This guide begins with an overview that uses examples to illustrate some key features of DataFerrett. After completing these examples, you will be familiar with four important features:

1. Choosing a dataset and variables
2. Adding variables to a DataBasket
3. Creating a Recode variable from a variable in your DataBasket
4. Creating a table in DataFerrett

The best way to become familiar with DataFerrett is to follow the steps in Examples One through Four. In these examples, you are looking for some basic information about families receiving cash for educational assistance. To start, access DataFerrett through its direct URL, <http://dataferrett.census.gov>. This is the main website for DataFerrett that contains general information, tutorials and a direct link to the application that is accessed directly from this page by clicking on the icon shown in Figure 0-1. From the application window, click on the icon in either Figure 0-2 or 0-3 to launch the DataFerrett program:



Figure 0-1: Link to DataFerrett Application



Figure 0-2: Link to Launch DataFerrett
(Production)

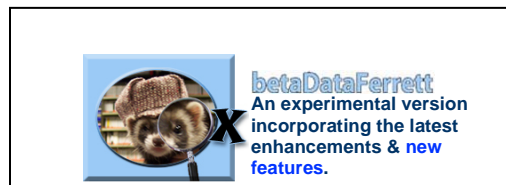


Figure 0-3: Link to Launch betaDataFerrett

NOTE: The DataFerrett RUN page gives you the option of choosing between the 'production' version (Figure 0-2) and the 'beta' version (Figure 0-3). The 'beta' version incorporates the newest features and enhancements before they become part of the 'production' version. The exercises throughout this guide use the 'production' version; however, you are free to use whichever version you prefer.

After clicking on the selected version, a login window will appear:

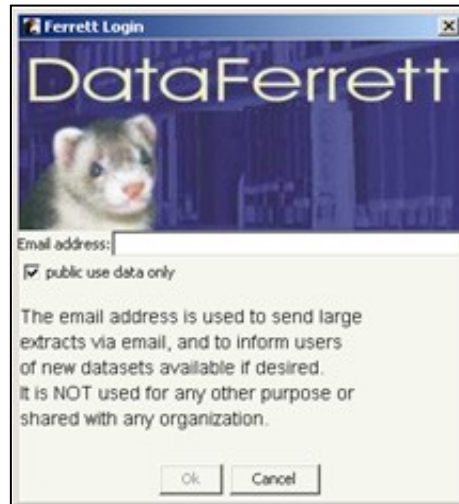


Figure 0-4: DataFerrett Login

Enter a valid email address in this window to access DataFerrett. By default, the box labeled “public use data only” will be checked. (For most users, this box must be checked.) Then click on the “OK” button (it will be active at that point). After entering your email address and clicking “OK,” you will arrive at the following screen:

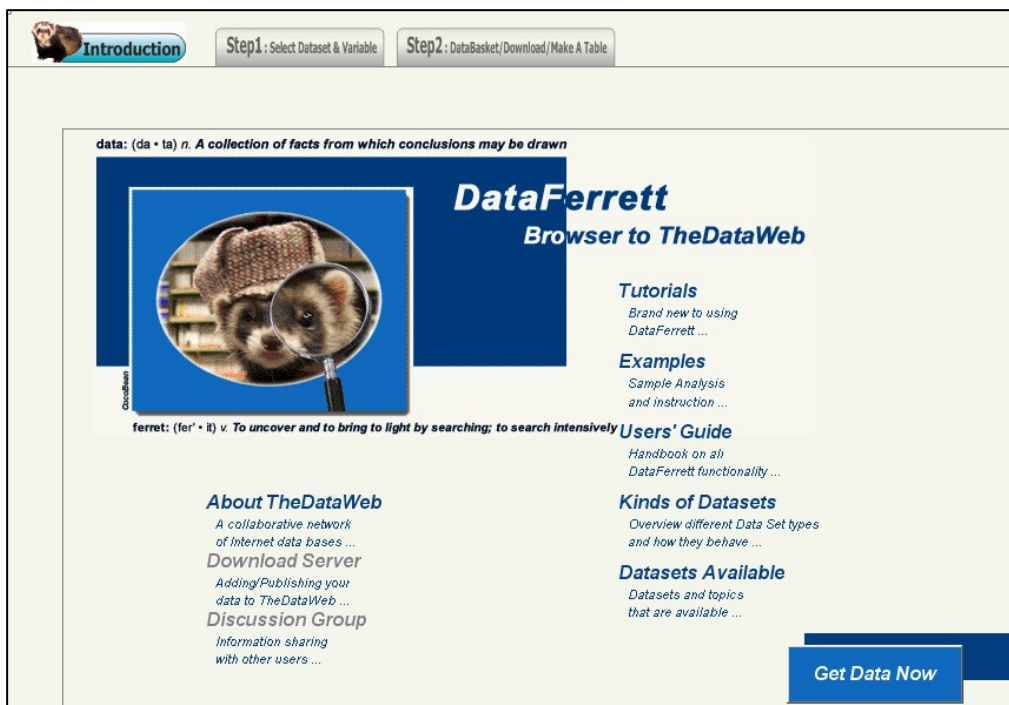


Figure 0-5: DataFerrett Opening Screen

To proceed, click on the “Get Data Now” button.

Example 1: Choose Dataset and Variables to View

The most recent information on the topic of families receiving cash assistance for education is available from the March 2010 Supplement of the Current Population Survey (CPS.) Figures 0-6 and 0-7 below demonstrate the sequence involved in selecting the dataset and choosing which group of variables to view. This example involves four user actions, as shown in Figures 0-6 and 0-7:

Example 1: Select Dataset and Choose Variable to View		
Steps	What to do	Why do it?
1	After arriving at the “Step 1” screen, click on the plus sign by Current Population Survey in the “Select Dataset(s) to Search.” Then select the “March Supplement” folder and highlight “Mar 2010”.	You must click on the plus sign in order to open the folder and select a dataset.
2.1 2.2 2.3	After highlighting Mar 2010, click on “View Variables” (2.1) from the fly out menu. The available topics will be shown to the right of the dataset list. Check the box in front of “Family Variables” (2.2) under “Select All Topics” then click on the magnifying glass icon (2.3).	March 2010 is an “instance” of the “CPS - March Supplement” dataset. Checking the “Family Variables” box restricts the number of variables to those you need. All variables contained within the box or boxes checked will be displayed after clicking on the magnifying glass.

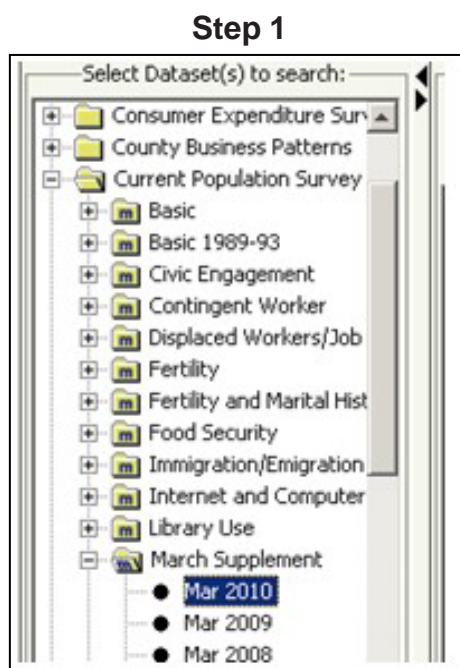


Figure 0-6: Selecting Dataset and Variables

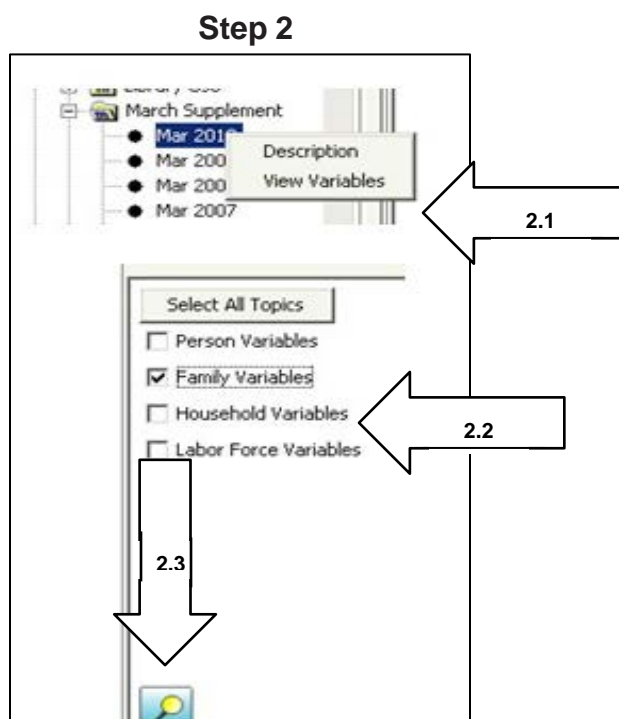


Figure 0-7: Select “Family Variables” from List of Topics

Example 2: Select Specific Variable and Add It to the DataBasket

After selecting the magnifying glass icon, the number and the list of available variables will be displayed. In this example there are 75 “Family Variables” that are available to place in your DataBasket, as shown in Figure 0-8. The following example demonstrates how to select a specific variable from that list to work with in “Step 2: DataBasket/Download/Make a Table.”

Example 2: Select Specific Variable and Add It to the DataBasket		
Steps	What to do	Why do it?
1.1 1.2	Find the variable containing the educational assistance amount for each family by highlighting FEDVAL (Education assistance amount - Family) (1.1), then click “Browse/Select Highlighted Variables” (1.2).	Clicking on “Browse/Select Highlighted Variables” takes you to the screen that places the variable in the DataBasket.
2.1 2.2	Click the box next to “Select” (2.1) then click “OK” (2.2).	This allows you to add the variable to the Data- Basket.
3	A confirmation window will pop-up. When it does, click “OK.”	This confirms the number of variables you have added to your DataBasket.
4	Click the “Step 2” tab.	This takes you to the “Step 2: DataBasket/ Download/Make a Table” screen where you can work with the variables you have chosen.

Step 1

Highlight the variables you are interested in				
75 Variables returned from search. 0 variables selected in DataBasket.				
<div style="text-align: right;"> <input type="button" value="Browse/Select Highlighted Variables"/> </div>				
Topic	Name	Availability	Variable Label	Mergeable
Family Variables	FFNGCAID	Mar 1992 - Mar 2010	Health Insurance,Medicaid-Family fungible value	No
Family Variables	FFNGCARE	Mar 1992 - Mar 2010	Health insurance,Medicare-Family fungible value	No
Family Variables	FALMVAL	Mar 1992 - current	Alimony amount - Family	No
Family Variables	FAMLIS	Mar 1992 - current	Poverty - ration family income/low-income level	No
Family Variables	FCSPVAL	Mar 1992 - current	Child support amount - Family	No
Family Variables	FDISVAL	Mar 1992 - current	Disability amount - Family	No
Family Variables	FDIVVAL	Mar 1992 - current	Dividends amount - Family	No
Family Variables	FEDVAL	Mar 1992 - current	Educational assistance amount - Family	No
Family Variables	FFINVAL	Mar 1992 - current	Financial assistance amount - Family	No
Family Variables	FFOODREQ	Mar 1992 - Mar 2010	Estimated fungible value of food stamps - Family	No
Family Variables	FFPOS	Mar 1992 - current	FF Family sequence indicator 1989 +	No
Family Variables	FFPOSOLD	Mar 1992 - current	FF Family sequence indicator 1988 <	No
Family Variables	FFRVAL	Mar 1992 - current	Farm income amount - Family	No
Family Variables	FHEADIDX	Mar 1992 - current	PPPOS value of family reference person	No

Figure 0-8: Highlight Variable

Step 2

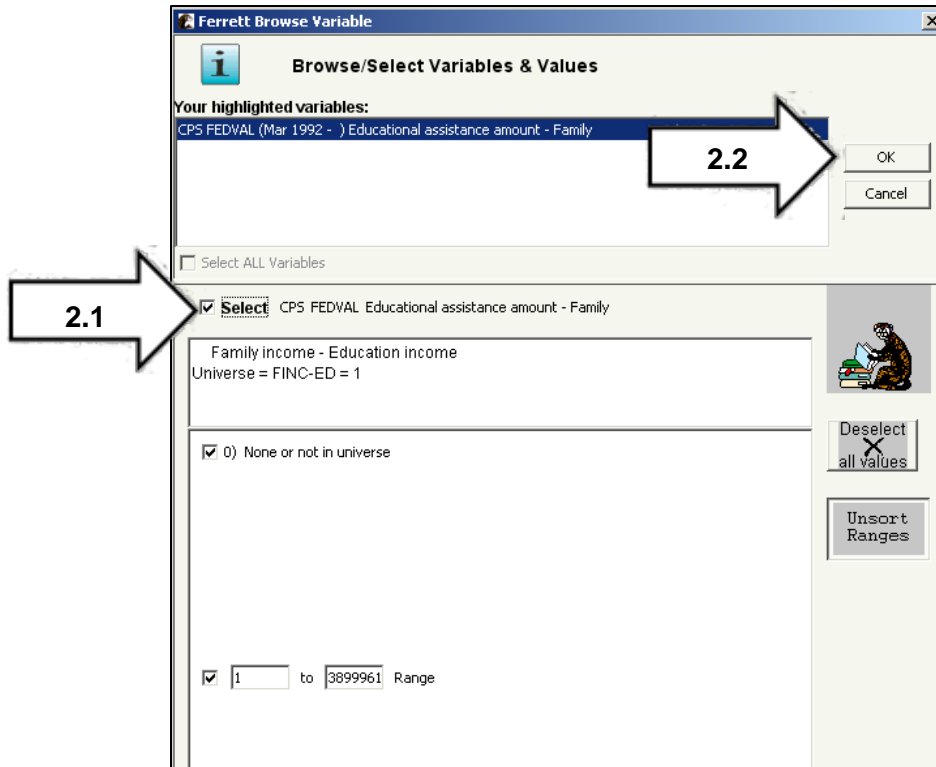


Figure 0-9: Pop-up Window for Choosing Variables

Step 3

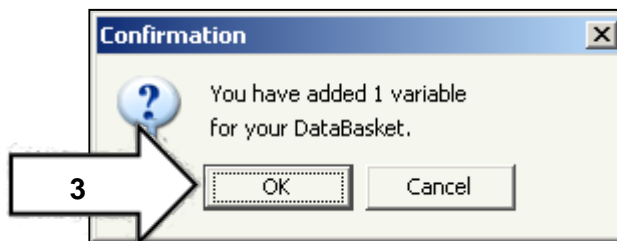


Figure 0-10: DataBasket Pop-Up Message

Step 4

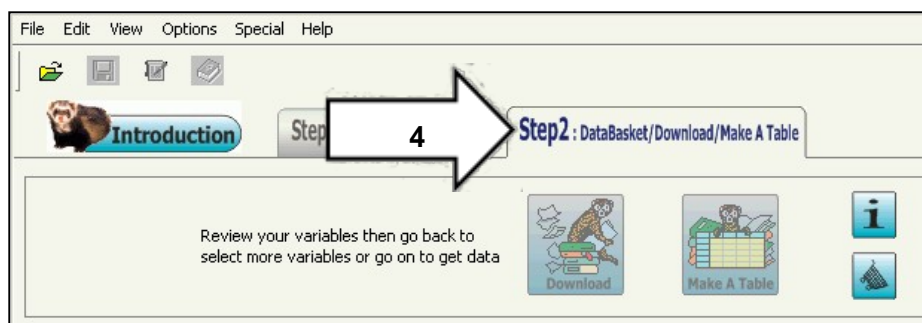


Figure 0-11: Go to "Step 2" Screen

NOTE: Information from individual DataFerrett sessions can be saved for future reference. See “Saving and Retrieving Information from a DataFerrett Session” for an explanation of Ferrett Tabulation Files and Ferrett Session Files.

Example 3: Create a “Recode” Variable

The variable added to the DataBasket in Example 2 will appear in the “Step 2: Databasket/Download/Make A Table” screen, where you can work with the variables in your DataBasket, including creating a “recode variable.”

Example 3: Create a “Recode” Variable		
Steps	What to do	Why do it?
1.1 1.2	Click on the variable label (1.1) and then the “Recode Variable” (1.2) button.	The “Recode Variable” feature allows you to group the values of a variable into categories that become the values of the new “recode variable.”
2.1 2.2 2.3	Type in “Ed assistance amt” in the blank line for the recode variable label (2.1), type in 10000 in the “continuous values” range (2.2) and click the “Recode” button (2.3).	Naming the recode variable makes it easier to identify. After user action 2.3, the new recode variable has the value “1.” This value covers all values in the FEDVAL variable in the range of 1 to 10,000.
3.1 3.2	To set a value range, go to “Within the range from 10001 through...” and enter 30000 (3.1). Click the “Recode” button (3.2). Continue creating recode values in increments of 20000 until you reach 90000 (defining the final range as 90001 through 120000).	Since the variable consists of a dollar amount with a very large range of possible values, the recode values will be set in increments of 20000.
4	When you are finished setting the recode values, click “OK.” Figure 0-16 shows the DataBasket after creating the recode variable (RECODE1).	This will add your recode label to the Databasket. You are now able to use the new variable you created in a table or to create a map (covered in later chapters). Figure 0-16 shows the range of values for FEDVAL corresponding to each of the values of the recode

Step 1

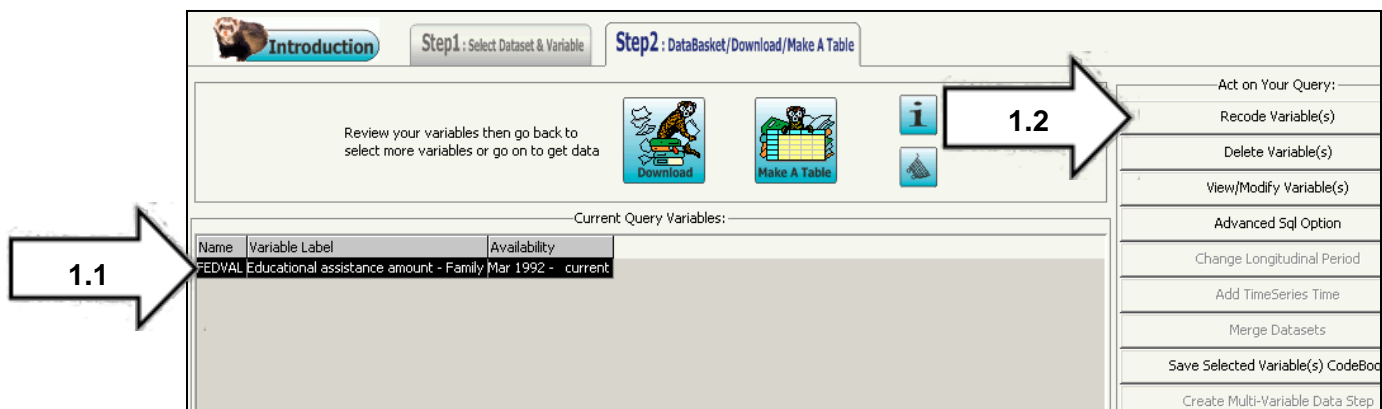


Figure 0-12: Selecting Recode Variables

Step 2

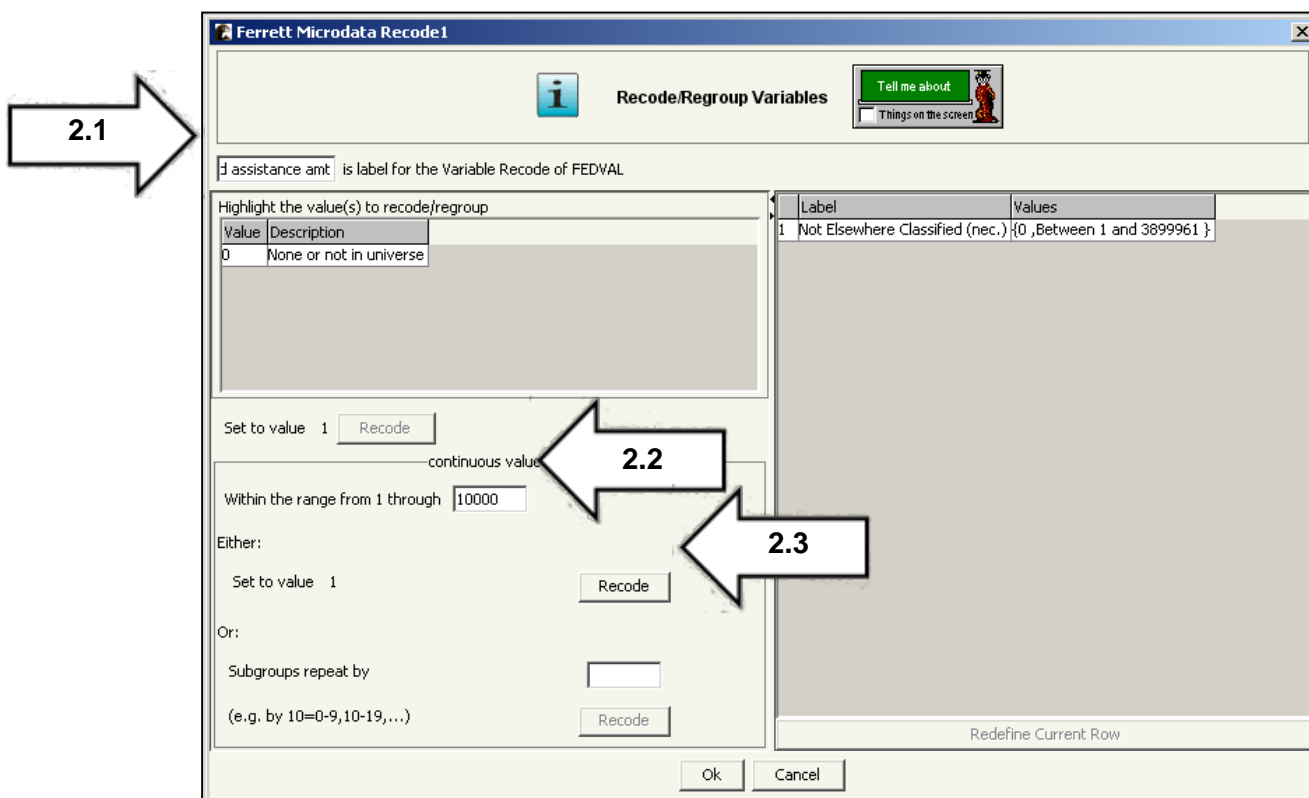


Figure 0-13: Begin Setting Recode

Step 3

Recode2 is label for the Variable Recode of FEDVAL

Highlight the value(s) to recode/ regroup

Value	Description
0	None or not in universe

Set to value 2 Recode

continuous values

Within the range from 10001 through 30000

Either:

Set to value 2 Recode

Or:

Subgroups repeat by Recode

(e.g. by 10=0-9,10-19,...)

Label	Values
1 Between 1 and 10000	{Between 1 and 10000}
2 Not Elsewhere Classified (nec.)	{0 ,Between 10001 and 3899961 }

Redefine Current Row

Ok Cancel

Figure 0-14: Recode/Regroup Variables Process

Step 4

Recode2 is label for the Variable Recode of FEDVAL

Highlight the value(s) to recode/ regroup

Value	Description
0	None or not in universe

Set to value 8 Recode

continuous values

Within the range from 10001 through 30000

Either:

Set to value 8 Recode

Or:

Subgroups repeat by Recode

(e.g. by 10=0-9,10-19,...)

Label	Values
1 Between 1 and 10000	{Between 1 and 10000}
2 Between 10001 and 30000	{Between 10001 and 30000}
3 Between 30001 and 50000	{Between 30001 and 50000}
4 Between 50001 and 70000	{Between 50001 and 70000}
5 Between 70001 and 90000	{Between 70001 and 90000}
6 Between 90001 and 120000	{Between 90001 and 120000}
7 Between 120001 and 3899961	{Between 120001 and 3899961}
8 Not Elsewhere Classified (nec.)	{0 }

Redefine Current Row

Ok Cancel


Figure 0-15: Recode Variables Added

Current Q		
Name	Variable Label	Availability
FEDVAL	Educational assistance amount - Family	Mar 1992 - current
RECODE1	Ed assistance amt	Mar 1992 - current

Figure 0-16: List of Variables

Example 4: Make a Data Table Using the Recoded Variable

For this example, you will be using the recode variable you created in Example Three.

Example 4: Make a Data Table Using the Recoded Variable		
Steps	What to do	Why do it?
1	<p>To create a data table with the variables in the DataBasket, click on the “Make a Table” icon in the “Step 2” screen.</p> 	This will open up an empty spreadsheet where you can place your variables to create a new table.
2	<p>After clicking on the “Make a Table” icon, the empty spreadsheet window contains a pop-up set of instructions superimposed on the window shown in Figure 0-17. Click on the “x” in the upper right-hand corner of the box to close it.</p>	The pop-up provides instructions on how to set up a table.
3.1 3.2	<p>To the right of the “Make a Table” screen, select the Recode variable (RECODE1) you created, as shown in Figure 0-18 (3.1), hold down the mouse button and drag the recode variable into the R2C1 cell (3.2) shown in Figure 0-19. The “?” symbol in column C2 is a placeholder for the numbers that will be displayed later.</p>	<p>Variables must be placed within a table before the data can be computed and the table can be made. The shorthand used in this example, R2C1, provides the row and column designation for the cell. Notice that the entry in R2 is labeled “Total RECODE1.” This row will contain the sum of the numbers for all eight values of RECODE1.</p>
4	<p>Once the table layout is displayed, click on the GO Get Data icon. Within seconds, the table is populated with the values for the RECODE1 variables for the U.S.</p>	<p>The table shows the distribution of educational assistance in dollars for families in the U.S. based on the March 2010 CPS Supplement using the value ranges as defined for the recode variable. Because the CPS is a sample survey, the numbers displayed are estimates for the whole population based on the sample.</p>
5	<p>Return to the “Step 2” screen to see the DataBasket with a new entry, the weight variable used for the tabulation.</p>	<p>When working with microdata based on a sample survey, DataFerrett can often determine the proper sample weight to use in a tabulation if only one type of weight is used; i.e., population or housing.</p>

The values shown in the table are weighted estimates. Note that you did not have to manipulate any data or complete any additional steps in order to get the weighted results. DataFerrett is able to determine which weights to use; a weighted number is the default result for any tabulation request where weights are present in the data (i.e., sample survey data.)

NOTE: You do not need to exit the “Make a Table” screen; you can minimize it and continue working in either the “Step 1” or “Step 2” screen if you wish to add variables or perform another recode. There are advantages to doing this – any update to your DataBasket will be present when you reactivate the “Make a Table” screen.

Step 2

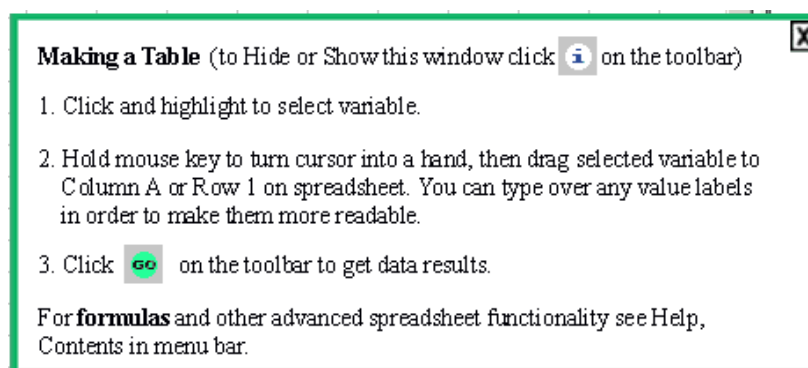


Figure 0-17: Make a Table Pop-up Window

Step 3

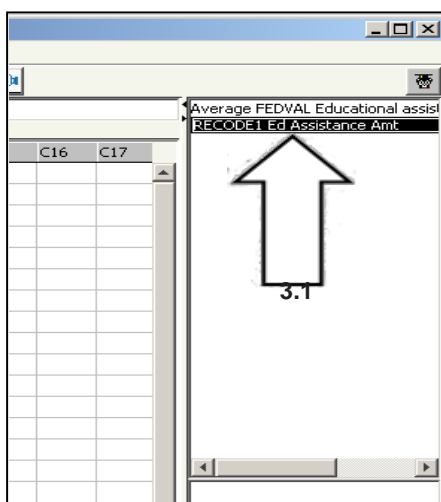


Figure 0-18: Recode Variable in “Make a Table” Screen

	C1	C2	C3
R1			
R2	Total RECODE1		?
R3	Between 1 and 10000		?
R4	Between 10001 and 30000		?
R5	Between 30001 and 50000		?
R6	Between 50001 and 70000		?
R7	Between 70001 and 90000		?
R8	Between 90001 and 120000		?
R9	Between 120001 and 3899961		?
R10	Not Elsewhere Classified (nec.)		?
R11			

Figure 0-19: Table Layout

Step 4

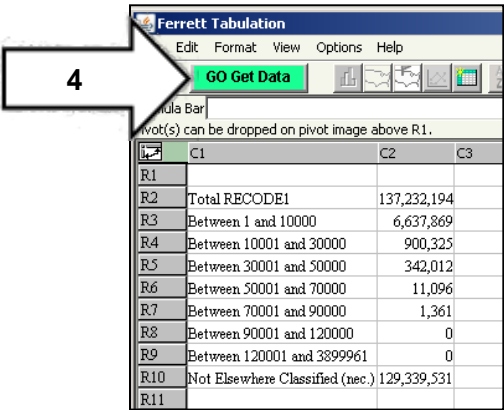


Figure 0-20: Data

Step 5

Name	Variable Label	Availability
FEDVAL	Educational assistance amount - Family	Mar 1992 - current
RECODE1	Ed assistance amt	Mar 1992 - current
FSUP_WGT	Weight, March supplement - Family	Mar 1992 - current

Figure 0-21: Weighted Results

Example 5: Saving and Retrieving Information From a DataFerrett Session

A table in DataFerrett can be saved for later use. A Ferrett Tabulation File (FTF) is created when you save the information that has been used to create a table, demonstrated in the following steps:

Saving & Retrieving Information from a DataFerrett Session		
Steps	What to do	Why do it?
1	Go to the “File” menu item on the upper left corner of the screen and select “Save As.”	This allows you to save it to whatever location you choose.
2	In the Save dialog box that pops up, choose the location where the file should be saved, name the file, and select “Ferrett Tabulation Files” for “Files of type.”	Saving data as a “Ferrett Tabulation File” allows you to retrieve data from a previous session and open it in the “Make a Table” screen.

3	If you are collecting variables that you would like to use later but have not yet laid out in a table, the file will be saved as a “Ferrett Session File” (FSF) containing only the variables currently in your DataBasket.	Creating and saving a “Ferrett Session File” allows you to view variables in your DataBasket without having to collect the data again. The FSF will also contain the definition of all recode variables in the
4	To retrieve your Ferrett Tabulation or Session File, go to the “File” menu and select “Open.” The table or selection of variables you have chosen appears in DataFerrett.	Once saved, Ferrett Tabulation or Session Files can be used for future DataFerrett applications.

Step 1

Ferrett Tabulation

File

Edit

Format

View

Options

Help

New

Open

Open in New Window

Save

Save As

Get Data

Debug

Print

Print Selected

Table Properties

Generate PDF

Graph

Map

Map Points

TimeSeries Graph

Rename

Exit

vot image above R1.

C2

Total FEDVAL

137,232,19

006,637,86

30000900,32

50000342,01

7000011,09

900001,36

120000

d 3899961

sified (rec.)129,339,53

Figure 0-22: Save Table Data

Step 2

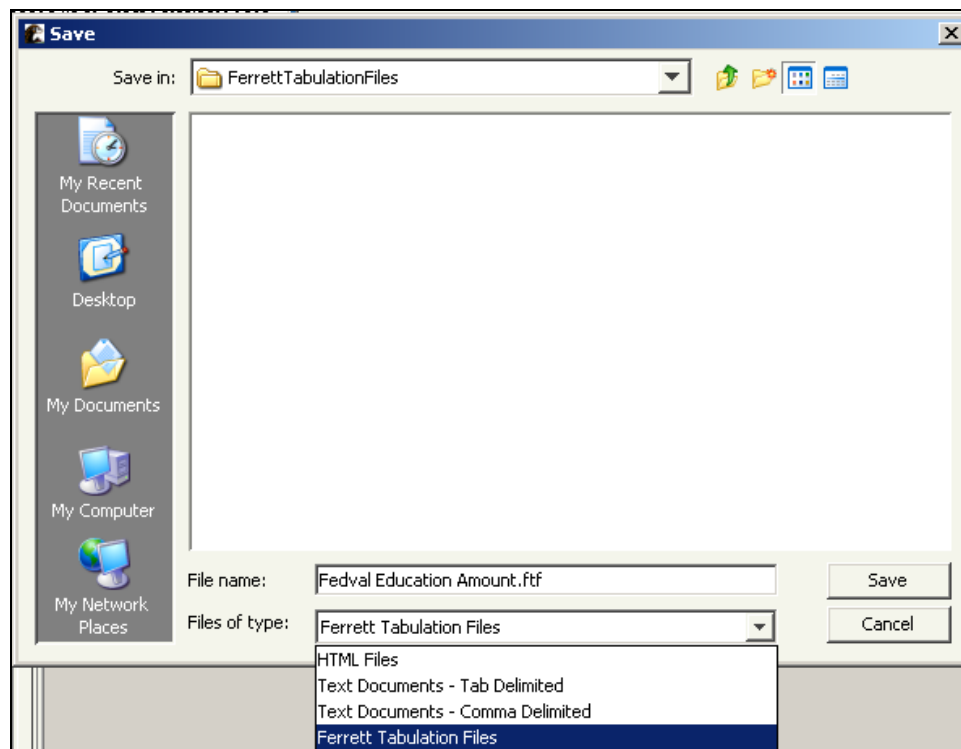


Figure 0-23: Save as “Ferret Tabulation File”

Step 3

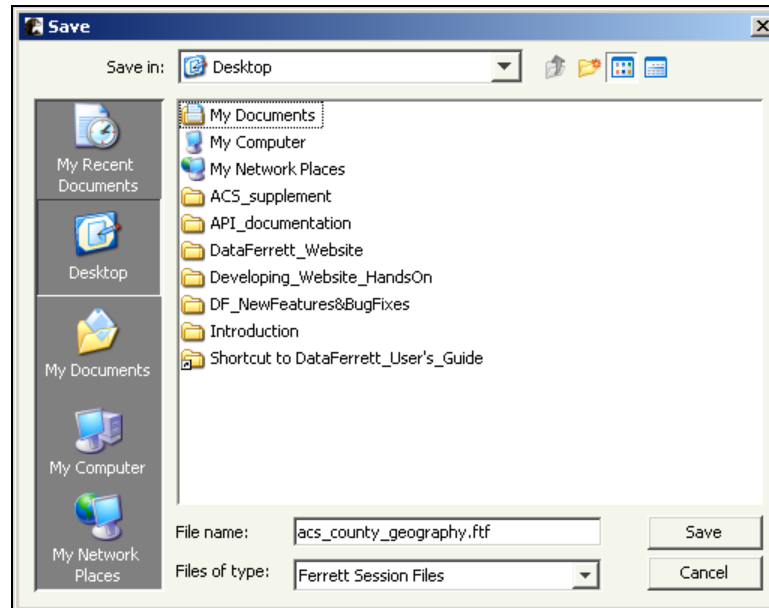


Figure 0-24: Save as “Ferret Session File”

Step 4

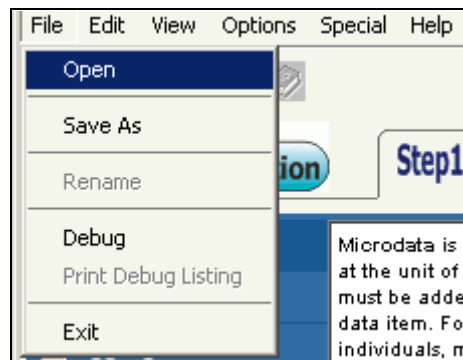


Figure 0-25: Open Saved File

CHAPTER 1: SELECTING VARIABLES & ADDING THEM TO THE DATA BASKET

A vast amount of data is stored within DataFerrett. This data comes from an array of sources such as:

- The American Community Survey – Updated Decennial data collected in rolling small area samples
- The Current Population Survey – Demographic, geographic and other characteristics of the US population and countries around the world
- County Business Patterns – Annual information about establishments with paid employees

Information from data sources such as these are referred to as “**datasets**” and organized into time periods such as one, three or five years. A specific time period is referred to as an “**instance**” and each instance contains characteristics of the dataset known as “**variables**.” Not all datasets have the same instances or variables, and not all instances within a dataset have the same variables. Individual variables have “**values**” that are unique to the variable, instance and dataset. Values associated with variables can be restricted to produce more focused and meaningful data analysis and presentation.

DataFerrett allows users to select variables and place them in a “**DataBasket**” where they can be saved, modified, or organized into presentation vehicles such as tables, graphs or maps. These data presentations provide the user with a visualization of the chosen data that greatly enhances analysis and understanding.

DataFerrett can only act on variables in the DataBasket. These variables must come from the same data source, i.e., a dataset and an instance within that dataset. The “Step 1: Select Dataset & Variables” window is the starting point for selecting datasets and variables within datasets. This chapter will describe how to use the features of “Step 1” to find variables and add them to the DataBasket.

The “Select Data Types” window in “Step 1” (Figure 1-1 below) displays the type of data in DataFerrett that is available for review and provides a description of each data type. You may choose to search individual or multiple data types by checking the box next to the data type title and clicking on the “Refresh Dataset List” button.

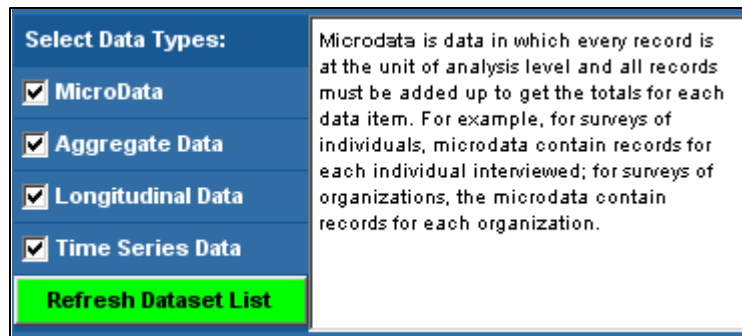


Figure 1-1: DataTypes

Once the data type(s) has been chosen, there are two ways to search variables and add them to the DataBasket:

1. Select a dataset and view the variables within that dataset.
2. Use the Search feature located at the top of the “Step 1” window.

Each of these methods is demonstrated in the following examples.

Example 1-1: View Variables by Choosing a Dataset

The following example demonstrates how to view variables when you know the dataset, variable and instance that you need. The 2009 Public Use Microdata Sample (PUMS) of the American Community Survey (ACS) is the specific instance searched by following the steps below. We begin at the “Step 1: Select Dataset & Variable” screen (Figure 1-2).

Example 1-1: Selecting Variables to View		
Step	What to do	Why do it?
1	In the “Select Datasets to Search” list, click the (+) sign next to the “American Community Survey” folder to see the ACS datasets.	Each folder in the dataset listing contains one or more datasets. All of the datasets in the selected folder are shown when the plus sign next to the data source title is clicked.
2	Click the (+) sign next to the “Public Use Microdata Sample” folder.	All available time periods, or instances, for which this dataset is available are shown when the plus sign next to the dataset title is clicked.
3.1 3.2	Click on the 2009 instance (3.1) from the bullet list within the dataset folder. A pop-up box with the choices “Description” and “View Variables” will appear. Click “View Variables” (3.2).	You can choose either to view a description of the dataset (i.e., the abstract) or to view the variables contained within the specified instance.

4.1 4.2	The variables included in the 2009 instance of the ACS PUMS are organized by topics on the “Select All Topics” screen. Check “Population” (4.1) and click “Search Variables” (4.2).	Topics are groupings of variables into categories to make it easier to find the variables of interest. The specific topics available are dependent on the dataset and the instance that have been chosen. To view variables, you can either click on “Select All Topics” (shows all variables) or select individual topics (shows variables for selected topics.)
5.1 5.2 5.3	Review the list of variables displayed and highlight the ones you want to examine more closely by holding down either the CTRL or the SHIFT key as you click on variables from the list. For this example, select “Decade of Entry” (5.1) and “Hispanic Recode” (5.2) and then click on “Browse/Select Highlighted Variables” (5.3).	Completion of this step opens the “Browse/Select Variables & Values” window where the values associated with your chosen variables can be seen and you can select which values to put in your DataBasket.

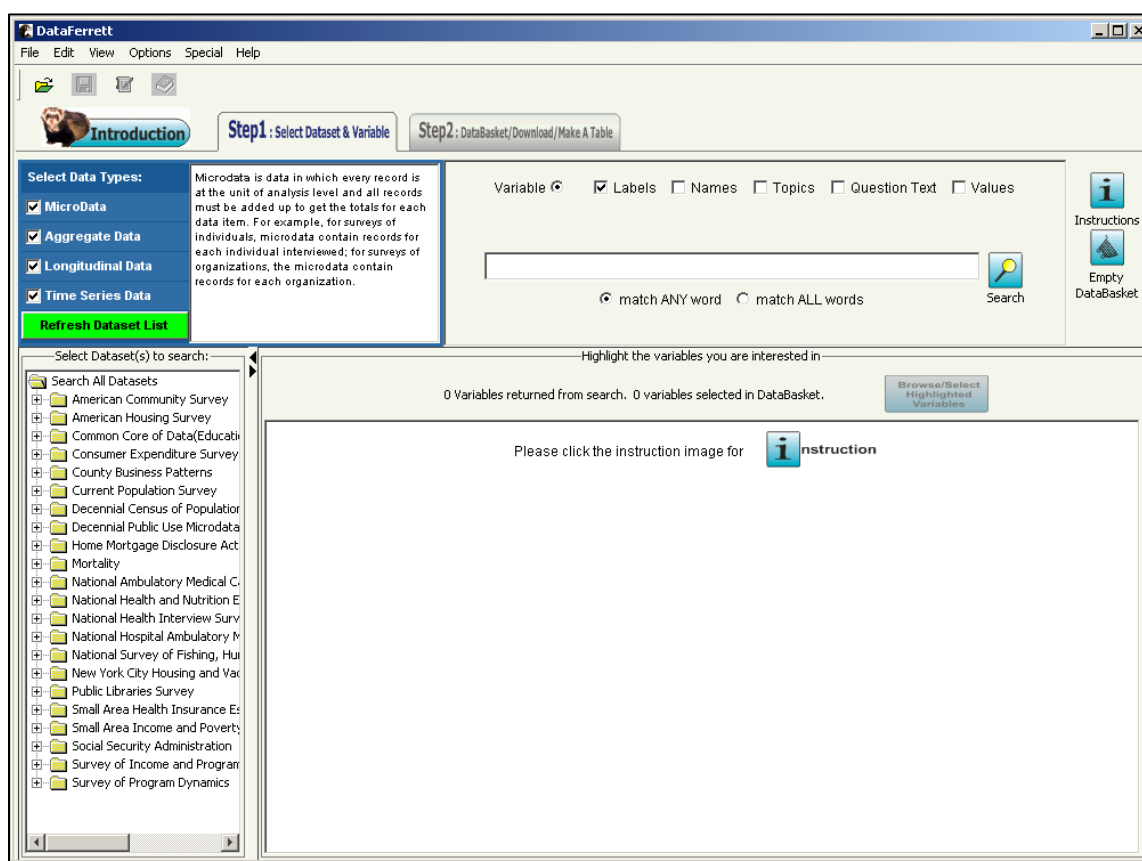


Figure 1-2: Select Dataset & Variable Screen



HINT: The “Select Dataset(s) to search” column can be resized by dragging either the arrows at the top of the column or the thick vertical bar to the right or the left.

Step 1

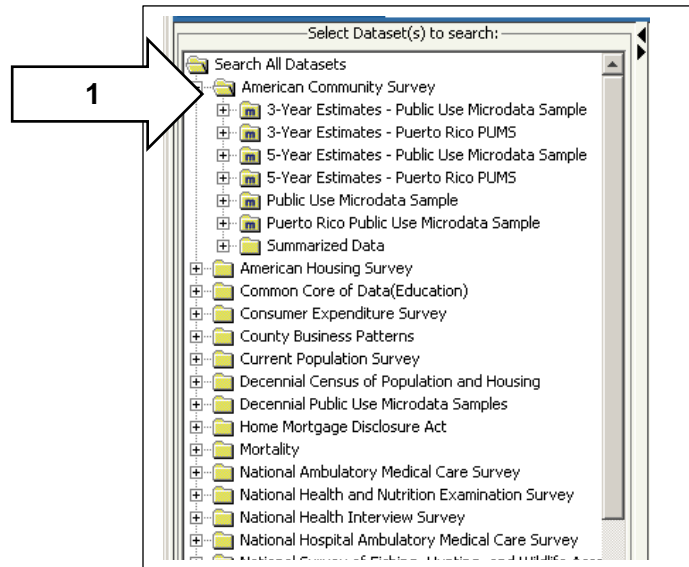


Figure 1-3: After Clicking on the “+” Sign for the American Community Survey Folder

Step 2

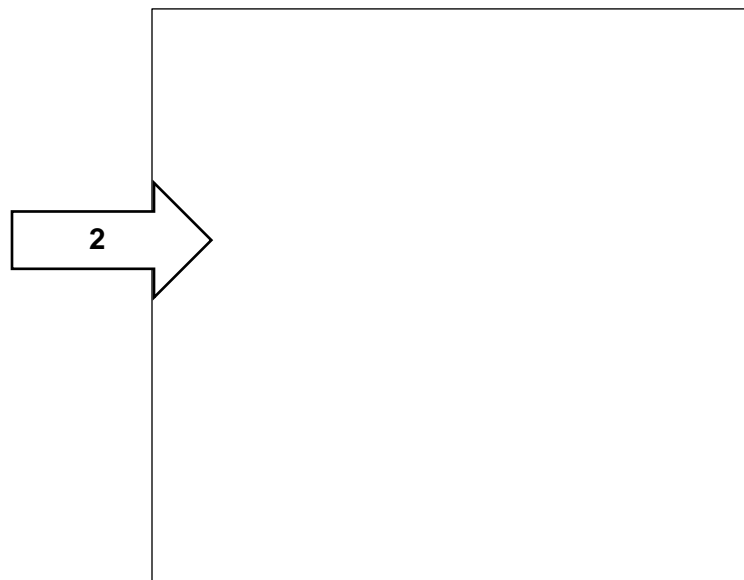


Figure 1-4: After Clicking on the “+” Sign for the Public Use Microdata Sample Subfolder

Step 3

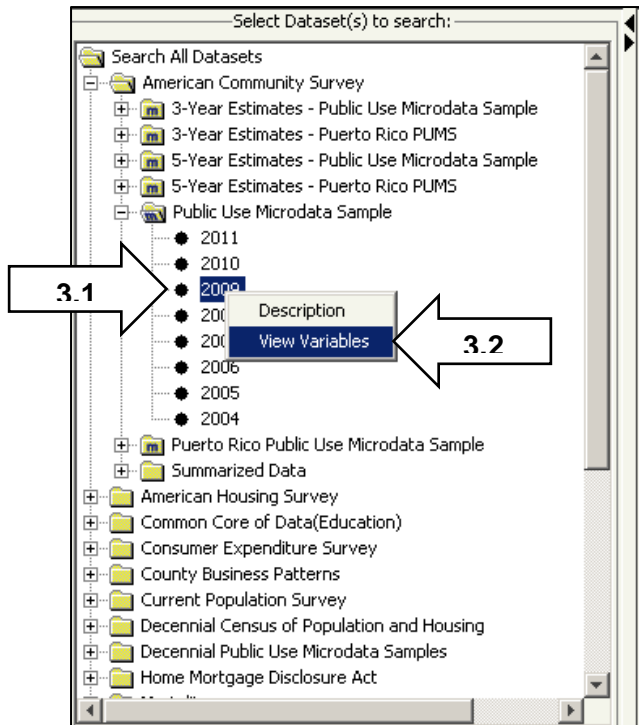


Figure 1-5: After Choosing the 2009 Instance of the ACS PUMS

Step 4

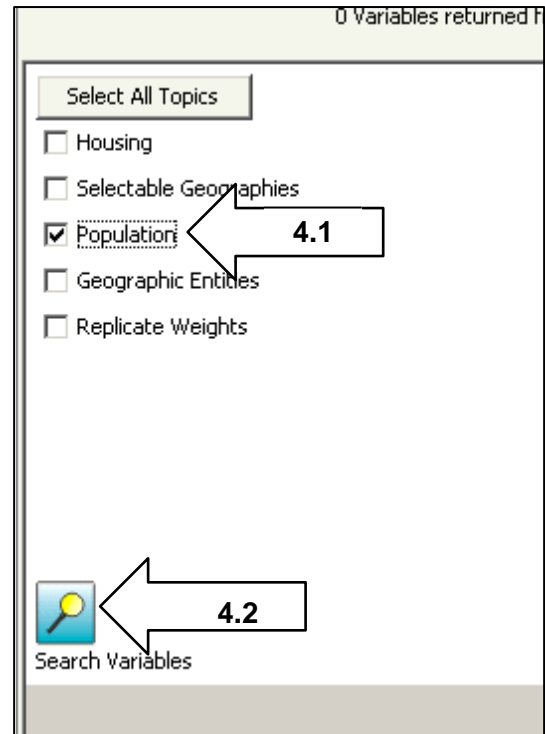


Figure 1-6: After Selecting Population Topic

Step 5

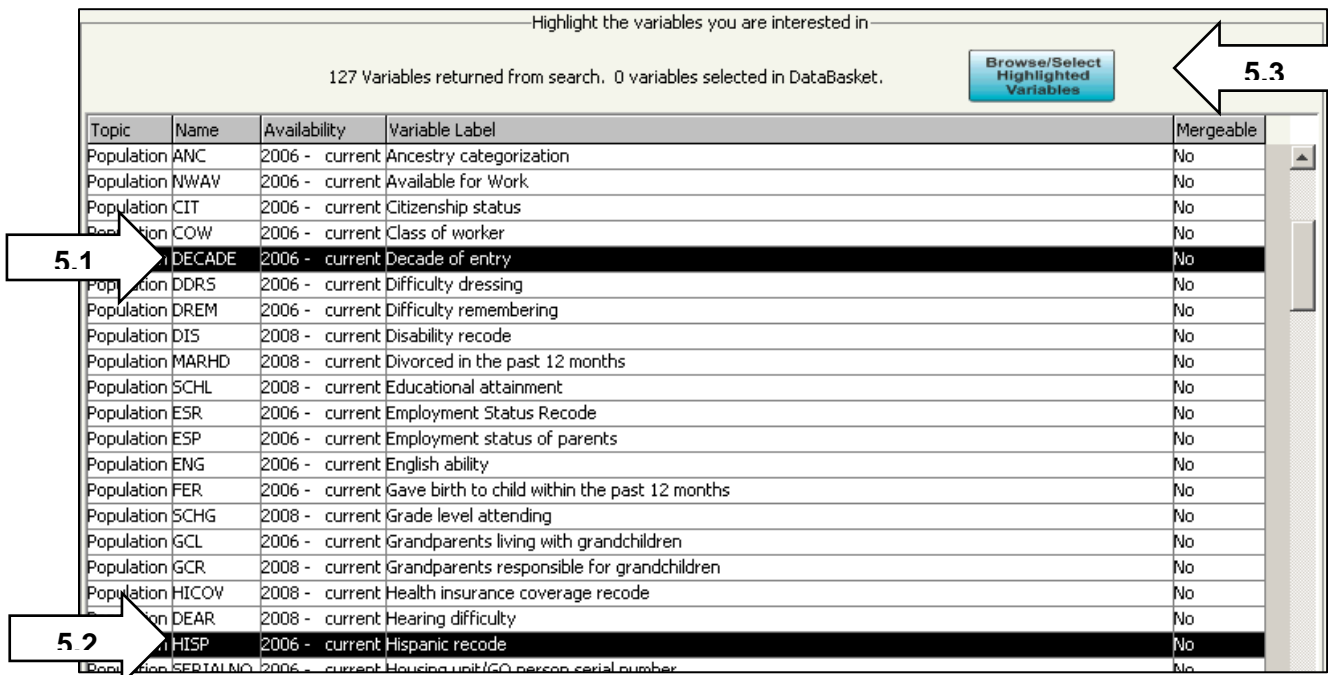


Figure 1-7: After Highlighting the Variables of Interest before Clicking “Browse/Select Highlighted Variables”



HINT: Columns in the table shown in Figure 1-7 can be sorted by clicking on the column title. Column widths can be changed by moving the dividing line between the column titles. Left-clicking the column title and dragging it to the desired location can also rearrange columns.

Descriptions of each column label in Figure 1-7 are provided in Table 1-1 below.

Column Label	Description
Topic	The high-level category for a variable, such as person, housing unit, labor force, business, etc. These categories will differ by dataset.
Name	The short mnemonic for the variable.
Availability	The time period during which the variable exists (existed) with the values and meanings as shown in the “Browse/Select Variables & Values” window.
Variable Label	Descriptive title of items existing within a dataset.
Mergeable	A “Yes” entry in this column indicates that the variable can be used as a key to link to a variable in a different dataset. Currently, only certain geography variables in certain datasets have a “Yes” value in this column.

Table 1-1: Definitions of Columns in Variable List

Example 1-2: View Variables by Searching

In addition to choosing specific variables within datasets, variables can be found by using the search function that is located at the top of the “Step1: Select Dataset & Variable” tab. The following examples explore several search scenarios: (1) searching all datasets with a single word, (2) searching within a single dataset with multiple instances by use of a single word, and (3) searching all datasets with a search phrase that contains multiple words. All examples begin at the “Step 1: Select Dataset & Variable” screen (Figure 1-2).

NOTE: At the top right of the “Step 1” tab is an area where variable search criteria such as labels, names, topics, question text and values can be entered by checking the box next to the desired search attribute. This is supplemented by a search box where a phrase or a series of words can be entered, along with the stipulation to “match ANY words” or “match ALL words.”

Example 1-2/Scenario 1: Searching All Datasets with a Single Word		
Step	What to do	Why do it?
1.1	Enter “tobacco” in the search box. Note that “Labels” and “match ANY word” are checked (1.1). Click “Search” (1.2).	“Tobacco” is the key word for this variable search.
1.2		DataFerrett defaults to “Labels” as the

		minimum search attribute and to “match ANY words” as the minimum search stipulation.
2.1 2.2	Highlight the variable name PEJ4 with the label “Stopped using the tobacco product in the last year” within the Current Population Survey (2.1) and click “Browse/Select Highlighted Variables” (2.2).	This step brings you to the “Browse/Select Variables & Values” screen that is required in order to add the variable to your DataBasket.
3.1 3.2	At the “Browse/Select Variables & Values” screen, check “Select” (3.1) and then “OK” (3.2).	This allows you to add the selected variable to your DataBasket. Alternatively, you may double-click the variable to add it to your DataBasket.
4	A pop-up window will appear that contains information on how many variables you have added to your DataBasket. If the displayed information is correct, click “OK.”	This step gives you the opportunity to confirm the number of variables that are placed in your DataBasket.
5	If the number of variables displayed in the confirmation pop-up window is incorrect, click “Cancel.”	Upon choosing “Cancel,” no variables are placed in your DataBasket and the selections of variables and values will need to be remade.

Step 1

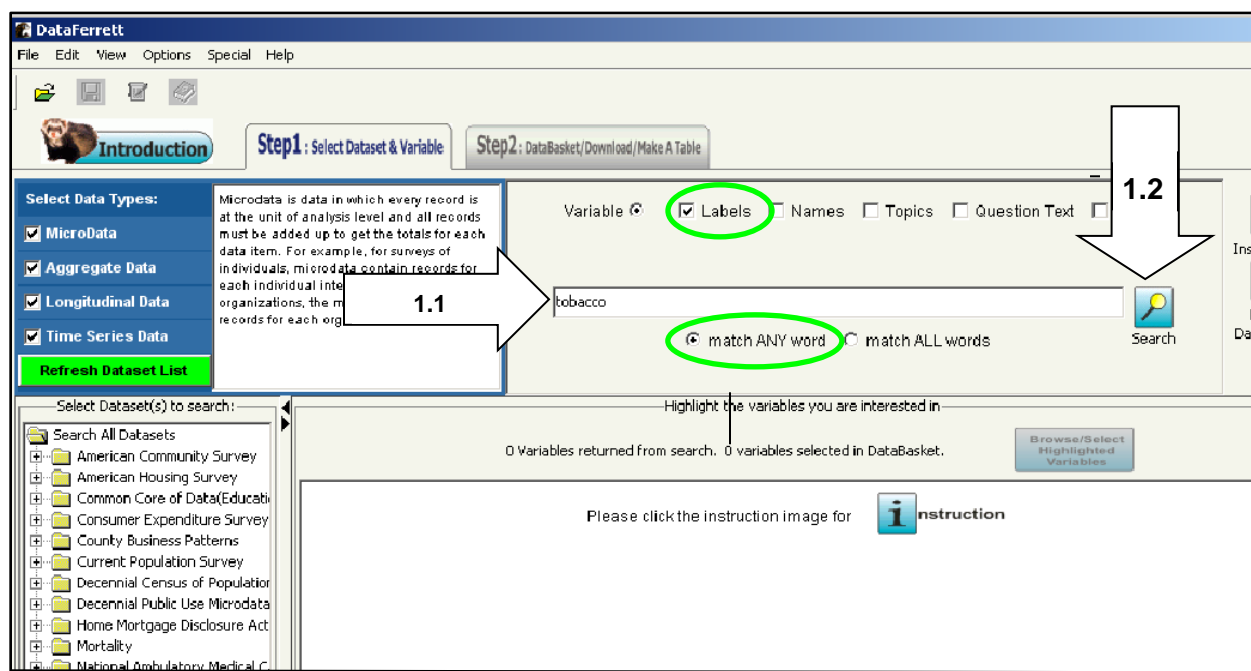


Figure 1-8: After Entering Single Word Search Term

Step 2

DataFerrett
File Edit View Options Special Help

Introduction Step1: Select Dataset & Variable Step2: DataBasket/Download/Make A Table

Select Data Types:
☒ MicroData
☒ Aggregate Data
☒ Longitudinal Data
☒ Time Series Data

Microdata is data in which every record is at the unit of analysis level and all records must be added up to get the totals for each data item. For example, for surveys of individuals, microdata contain records for each individual interviewed; for surveys of organizations, the microdata contain records for each organization.

Variable ☒ Labels ☐ Names ☐ Topics ☐ Question Text ☐ Values

tobacco

☒ match ANY word ☐ match ALL words

Select Dataset(s) to search: Highlight the variables you are interested in

118 Variables returned from search. 0 variables selected

2.1

2.2

Dataset	Name	Availability	Variable Label
CPS (Tobacco Use)	to Use Supplement Variables	May 2006 - Jan 2007	Number of days used chewing tobacco in last m
CPS (Tobacco Use)	to Use Supplement Variables	May 2010 - May 2010	Number of days used smokeless tobacco per m
CPS (Tobacco Use)	to Use Supplement Variables	Sep 1995 - May 1996	Smoking-advertising tobacco products should b
CPS (Tobacco Use)	to Use Supplement Variables	Sep 1995 - May 1996	Smoking-easy for minors to buy tobacco produc
CPS (Tobacco Use)	to Use Supplement Variables	Sep 1995 - May 1996	Smoking-ever used chewing tobacco on a regul
CPS (Tobacco Use)	to Use Supplement Variables	Sep 1995 - May 1996	Smoking-should free samples by tobacco compa
CPS (Tobacco Use)	to Use Supplement Variables	Sep 1995 - May 1996	Smoking-used pipes, cigars, chewing tobacco, c
CPS (Tobacco Use)	Tobacco Use Supplement Variables	May 2006 - Jan 2007	Stopped using the tobacco product in past year
CPS (Tobacco Use)	Tobacco Use Supplement Variables	May 2010 - May 2010	Switch to smokeless tobacco

Figure 1-9: Selected Variable

Step 3

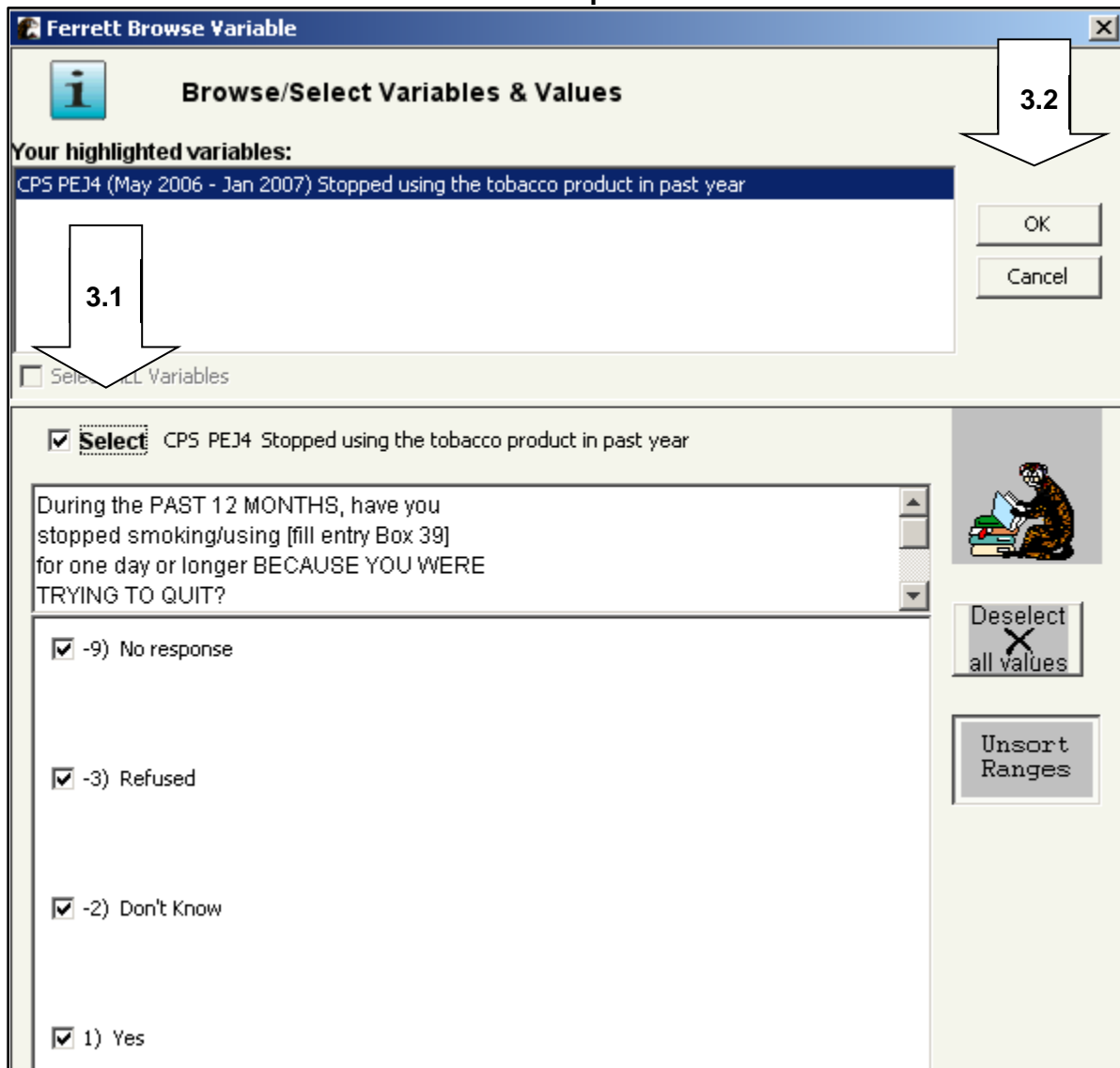


Figure 1-10: Variables Selected for DataBasket

Step 4/5

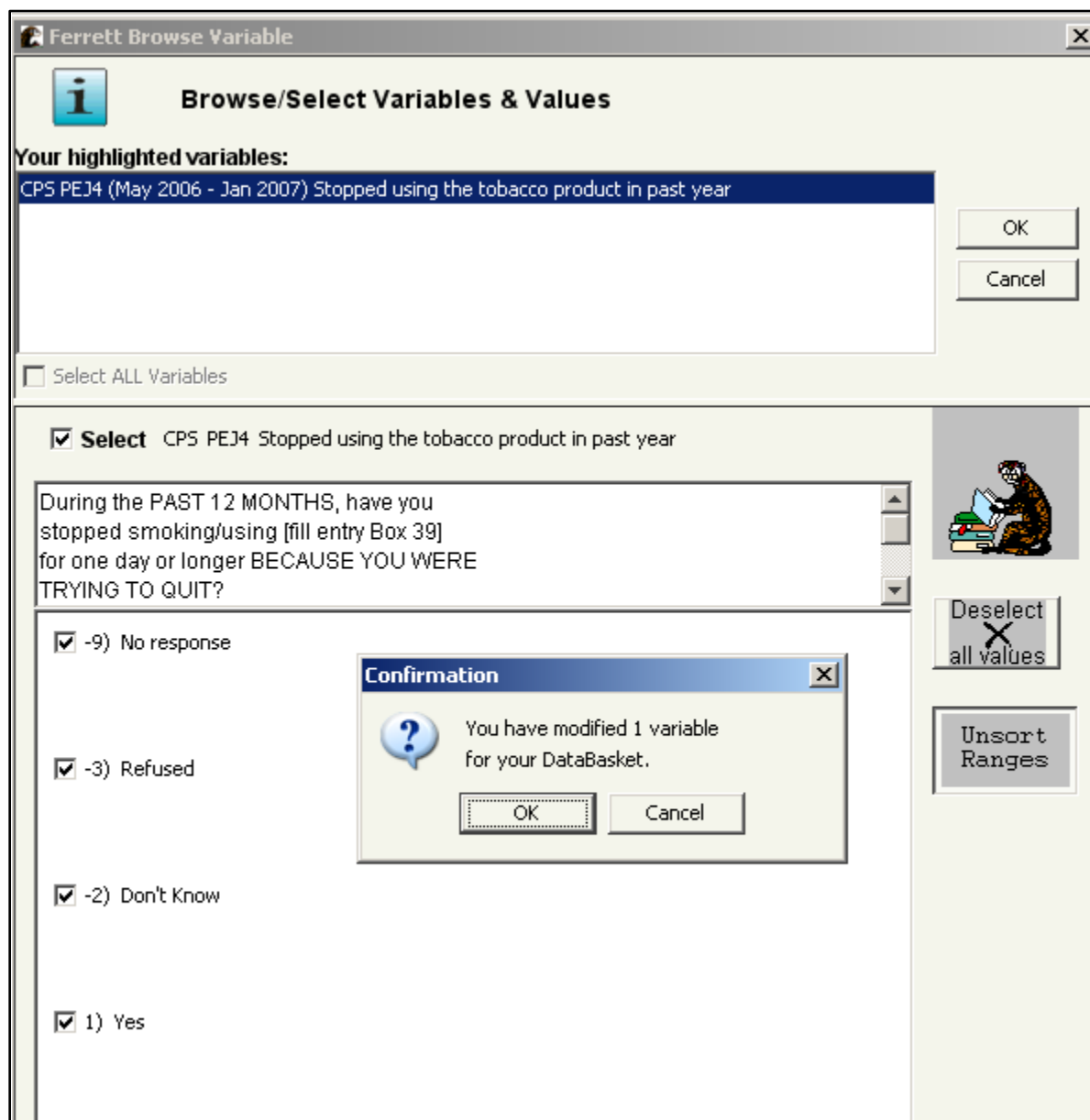


Figure 1-11: Confirmation of Number of Variables in DataBasket

Example 1-2/Scenario 2: Searching a Single Dataset/Multiple Instances with a Single Word		
Step	What to do	Why do it?
1.1 1.2 1.3 1.4	Enter “A_FTLF” in the search box (1.1), uncheck the “Labels” box and check the “Names” box (1.2). Allow “match ANY word” (1.3) to remain checked and click “Search” (1.4).	“A_FTLF” is the key word for this variable search. The automatically checked “Labels” box must be unchecked because this search is only concerned with the actual name of the variable, not its description.
2.1 2.2	Highlight the “A_FTLF” variable from the March supplement of the Current Population Survey (2.1) and click “Browse/Select Highlighted Variables” (2.2).	This step brings you to the “Browse/Select Variables & Values” screen that is required in order to add the variable to your DataBasket.
3.1 3.2	At the “Browse/Select Variables & Values” screen, check “Select” (3.1) and then “OK” (3.2).	This allows you to add the selected variable to your DataBasket.
4	A pop-up window will appear that contains information on how many variables you have added to your DataBasket. If the displayed information is correct, click “OK.”	This step gives you the opportunity to confirm the number of variables that are placed in your DataBasket.
5	If the number of variables displayed in the confirmation pop-up window is incorrect, click “Cancel.”	Upon choosing “Cancel,” no variables are placed in your DataBasket and the selections of variables and values will need to be remade.

Chapter 1: Selecting Variables & Adding Them to the DataBasket

Step 1

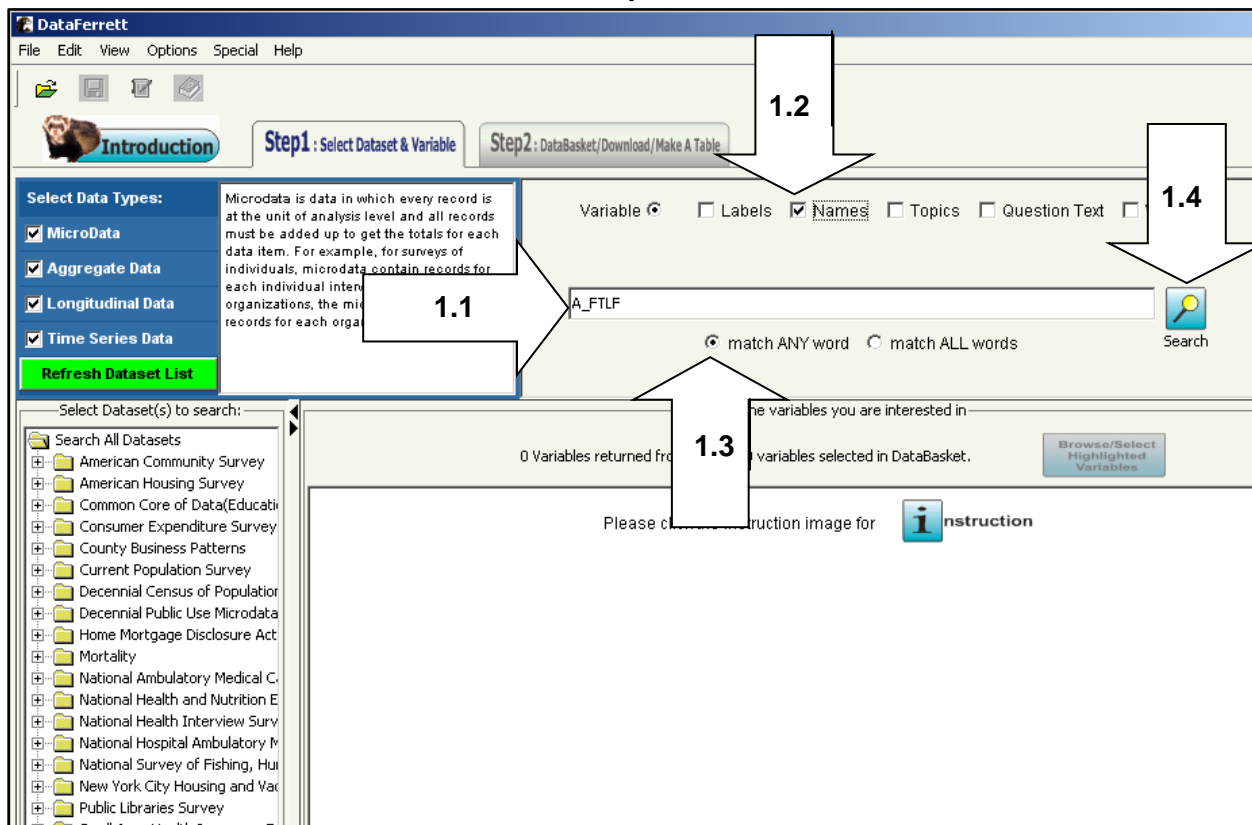


Figure 1-12: Entering Search Criteria

Step 2

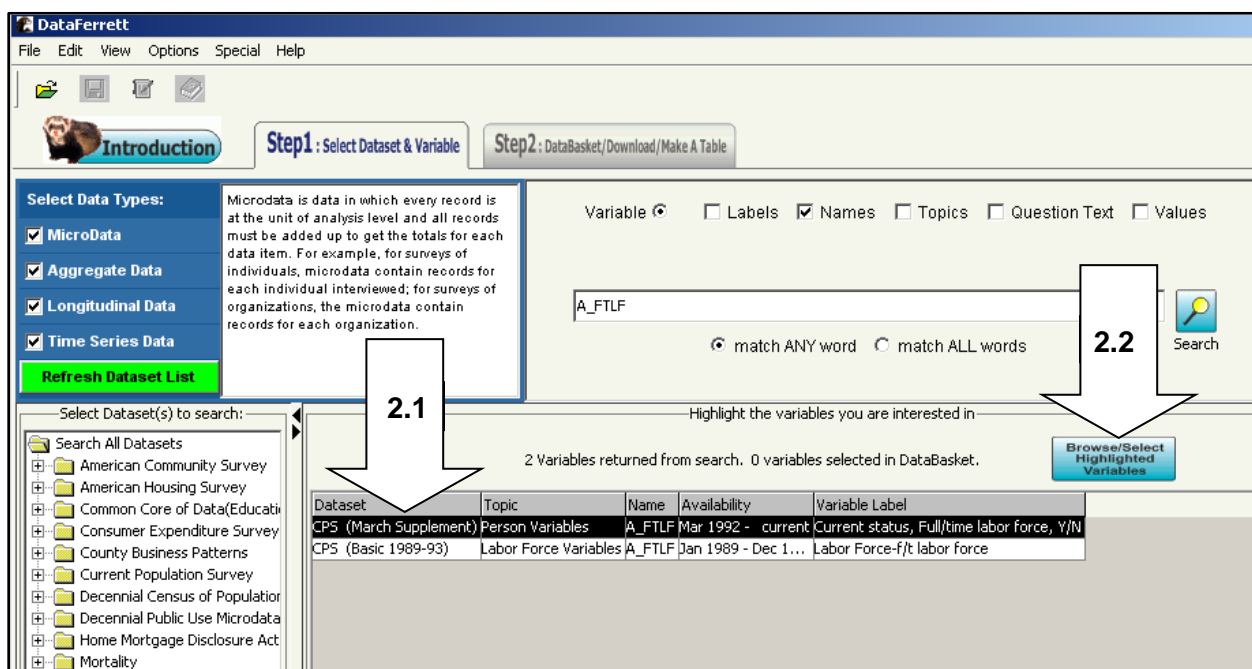


Figure 1-13: Selected Variable

Step 3

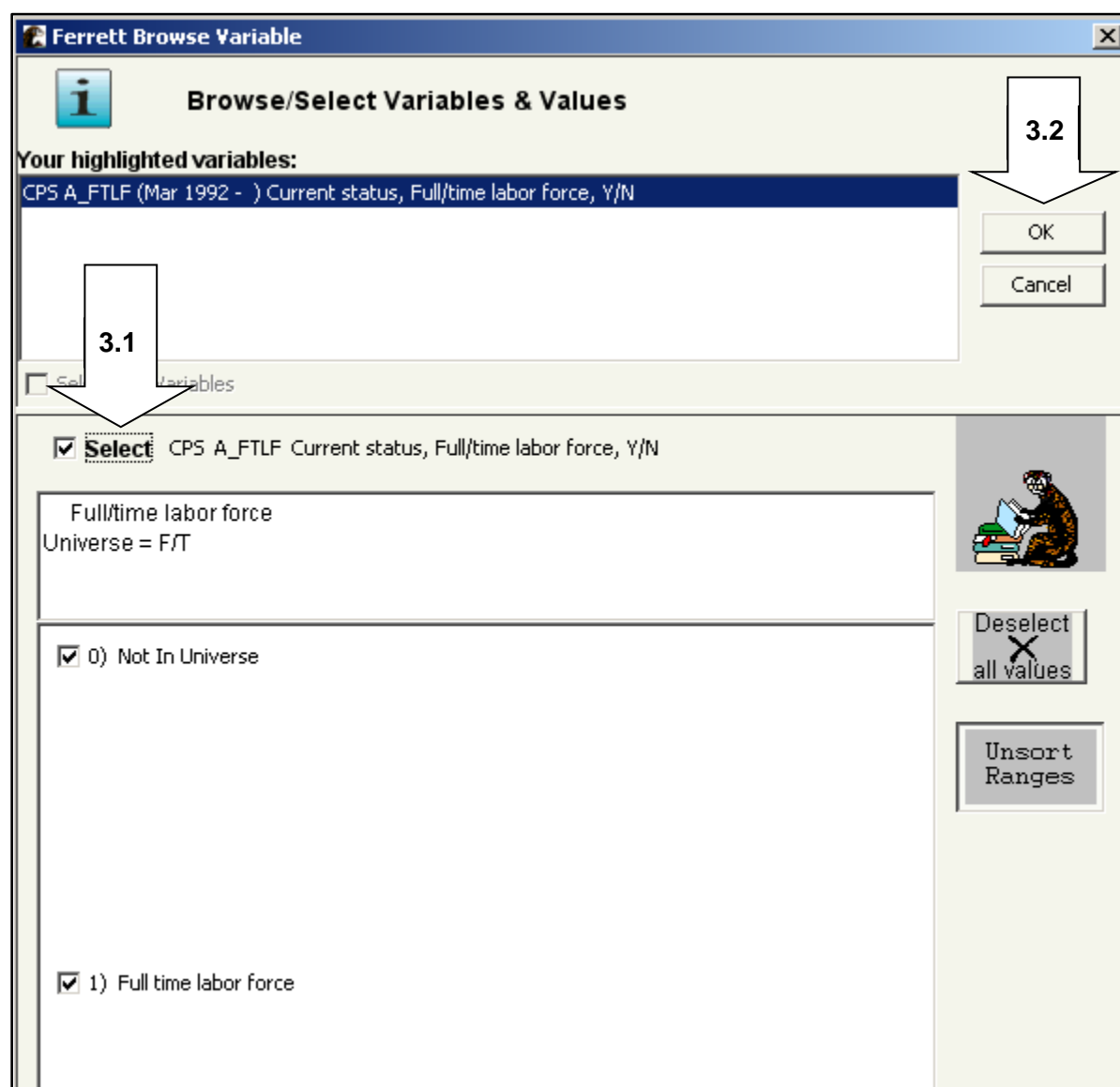


Figure 1-14: Variables Selected for DataBasket

Step 4/5

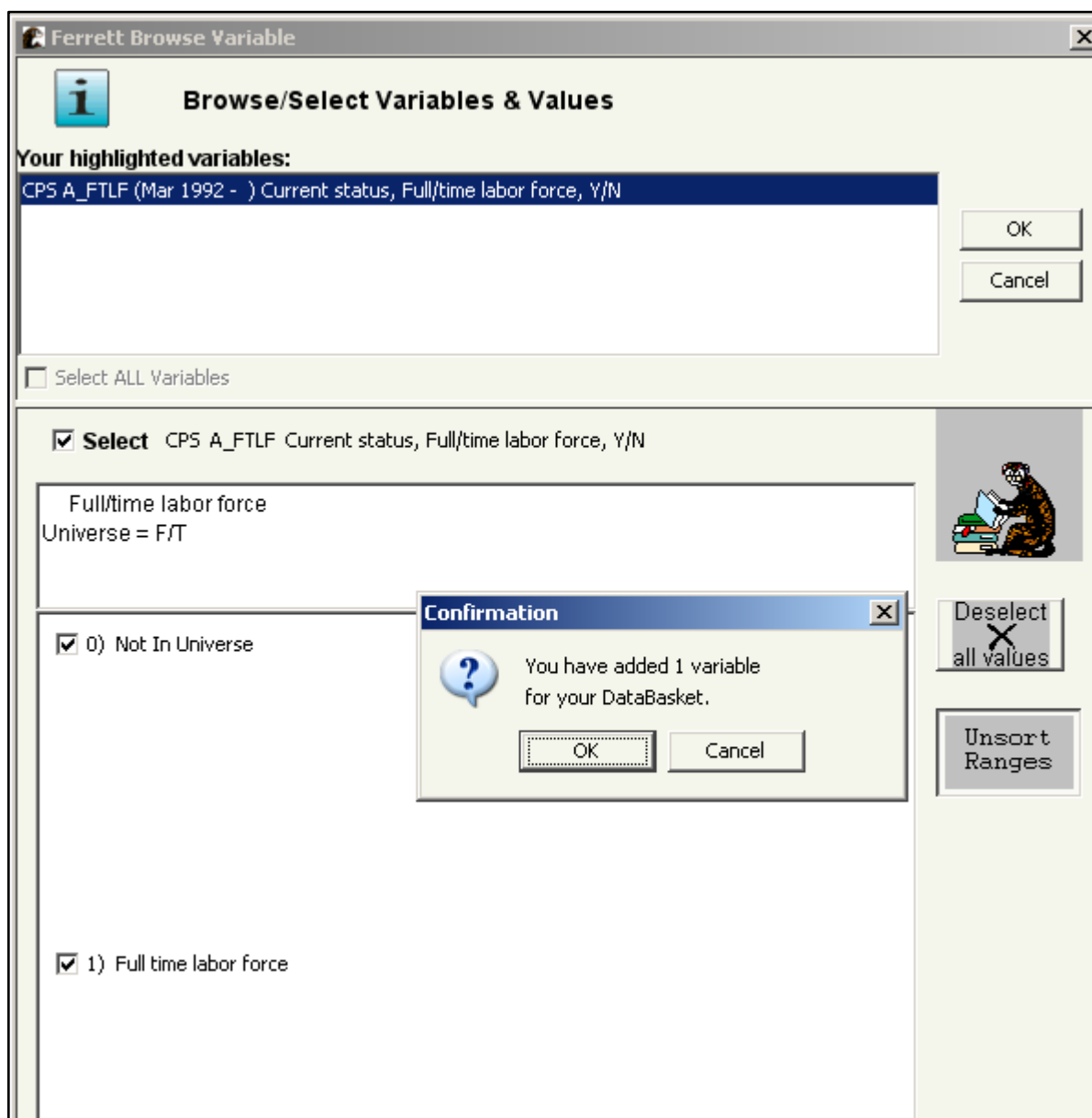


Figure 1-15: Confirmation of Number of Variables in DataBasket

Example 1-2/Scenario 3: Searching All Datasets with Multiple Words		
Step	What to do	Why do it?
1.1 1.2 1.3	Enter “home school” in the search box (1.1). Uncheck “match ANY words” and check “match ALL words” (1.2). Allow “Labels” to remain checked and click “Search” (1.3).	“home school” are the key words for this variable search. The automatically checked “match ANY words” box must be unchecked because this search requires that ALL key words match when returning variables.
2.1 2.2	Highlight the “PES62” variable from the Current Population Survey (2.1) and click “Browse/Select Highlighted Variables” (2.2).	This step brings you to the “Browse/Select Variables & Values” screen that is required in order to add the variable to your DataBasket.
3.1 3.2	At the “Browse/Select Variables & Values” screen, check “Select” (3.1) and then “OK” (3.2).	This allows you to add the selected variable to your DataBasket.
4	A pop-up window will appear that contains information on how many variables you have added to your DataBasket. If the displayed information is correct, click “OK.”	This step gives you the opportunity to confirm the number of variables that are placed in your DataBasket.
5	If the number of variables displayed in the confirmation pop-up window is incorrect, click “Cancel.”	Upon choosing “Cancel,” no variables are placed in your DataBasket and the selections of variables and values will need to be remade.

Step 1

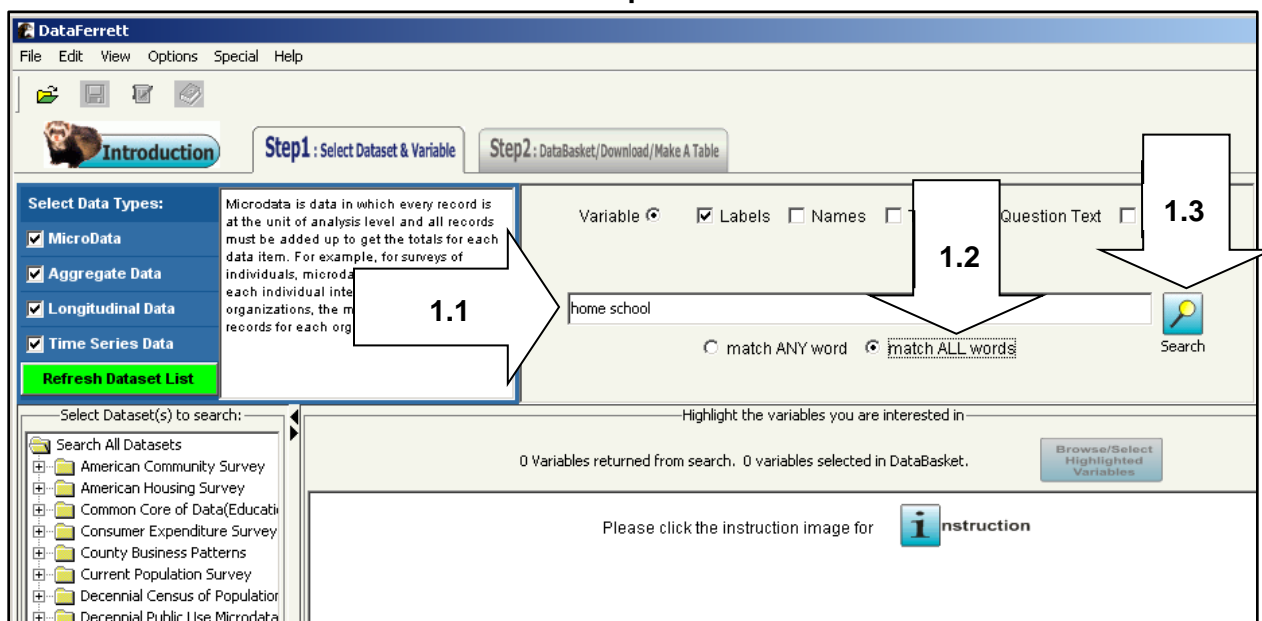


Figure 1-16: Entering Search Criteria

Step 2

DataFerrett
File Edit View Options Special Help

Step1 : Select Dataset & Variable **Step2 : DataBasket/Download/Make A Table**

Select Data Types:

- ☒ MicroData
- ☒ Aggregate Data
- ☒ Longitudinal Data
- ☒ Time Series Data
-

Microdata is data in which every record is at the unit of analysis level and all records must be added up to get the totals for each data item. For example, for surveys of individuals, microdata contain records for each individual interviewed; for surveys of organizations, the microdata contain records for each organization.

Variable ☒ Labels ☐ Names ☐ Topics ☐ Question Text ☐ Values

home school

☐ match ANY word ☒ match ALL words

2.2

Search

Select Dataset(s) to search:

- Search All Datasets
- American Community Survey
- American Housing Survey
- Common Core of Data(Educational)
- Consumer Expenditure Survey
- County Business Patterns
- Current Population Survey
- Decennial Census of Population
- Decennial Public Use Microdata
- Home Mortgage Disclosure Act
- Mortality
- National Ambulatory Medical C...
- National Health and Nutrition E...
- National Health Interview Surv...
- National Hospital Ambulatory M...
- National Survey of Fishing, Hui...
- New York City Housing and Vac...

Highlight the variables you are interested in

17 Variables returned from search. 0 variables selected in DataBasket.

2.1

Dataset	Topic	Name	Availability	Variable Label
SIPP 1996 Panel (Topic...	Adult Well Being Topical Module Variables	EAPHOMS	Wave 8 - Wave 8	AW: Children attend home school
SIPP 2001 Panel (Topic...	Adult Well Being Topical Module Variables	EAPHOMS	Wave 8 - Wave 8	AW: Children attend home school
SIPP 2004 Panel (Topic...	Adult Well Being Topical Module Variables	EAPHOMS	Wave 5 - Wave 5	AW: Children attend home school
SIPP 2008 Panel (Topic...	Adult Well Being Topical Module Variables	EAPHOMS	Wave 6 - Wave 6	AW: Children attend home school
SIPP 2008 Panel (Topic...	Adult Well Being Topical Module Variables	EAPHOMS	Wave 9 - Wave 9	AW: Children attend home school
CPS (Internet and Tele...	Computer Ownership/Internet Supplem...	PESAWAY1	Oct 1997 - Oct 1997	Computer-student reside at home or school
CPS (Internet and Tele...	Computer Ownership/Internet Supplem...	PESCU6L	Oct 1997 - Oct 1997	Computer-home use for school assignments,y/n
CPS (Internet and Tele...	Computer Ownership/Internet Supplem...	PESCU4F	Oct 1997 - Oct 1997	Computer-(child)use at home for school assignmen
CPS (Internet and Tele...	Computer Ownership/Internet Supplem...	PESCCU4I	Oct 1997 - Oct 1997	Computer-(child)use at home to connect to school
CPS (School Enrollment)	School Enrollment Supplement Variables	PES62	Oct 1994 - Oct 1994	Home schooled-not attending regular school-reaso
CPS (School Enrollment)	School Enrollment Supplement Variables	PES69	Oct 1994 - Oct 1994	Home schooled-not attending school-reason
CPS (School Enrollment)	School Enrollment Supplement Variables	PES72	Oct 1994 - Oct 1994	Home schooled-school type attending
CPS (School Enrollment)	School Enrollment Supplement Variables	PES72SP	Oct 1994 - Oct 1994	Home schooled-school type(other)-specify

Browse/Select Highlighted Variables

Figure 1-17: Selected Variable

Step 3

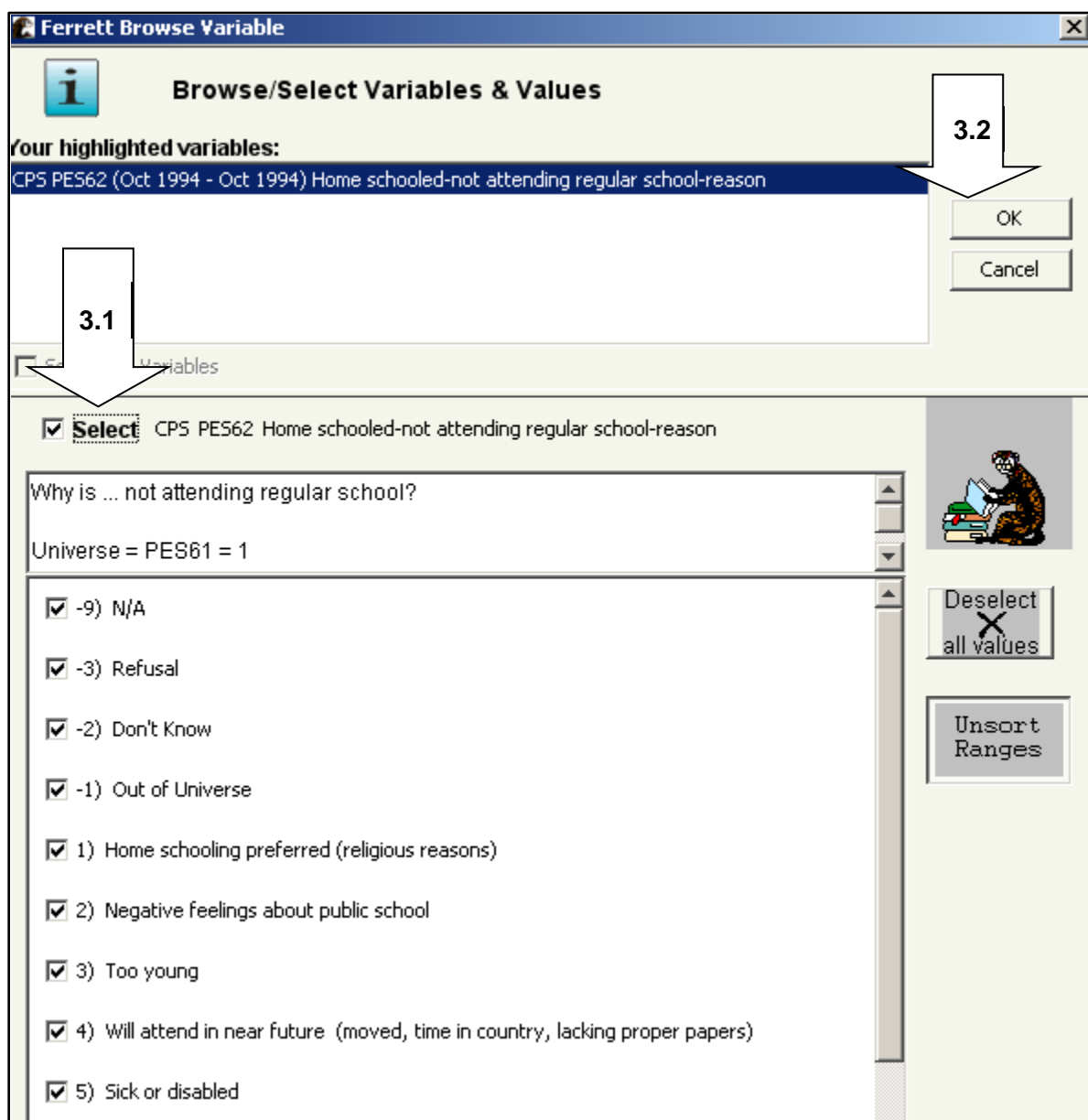


Figure 1-18: Variables Selected for DataBasket

Step 4/5

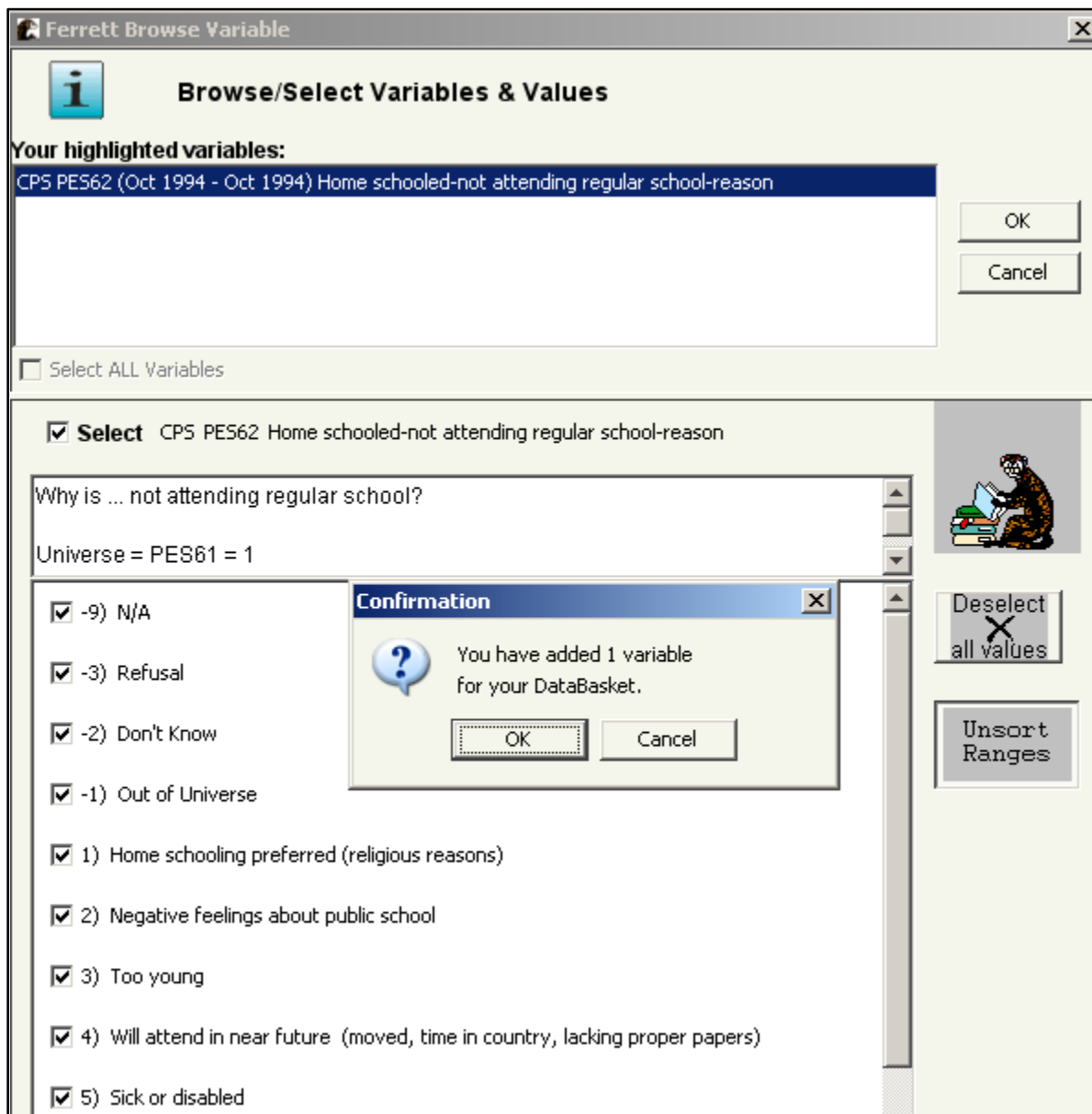


Figure 1-19: Confirmation of Number of Variables in DataBasket

Adding Variables to the DataBasket

After you have chosen variables, you will need to review the values associated with them to decide if they are appropriate for your current work. This review occurs on the “Browse/Select Variables & Values” screen where you can choose to include all or selected values for individual or multiple variables. The examples below demonstrate how to execute the options for adding variables and their values to the DataBasket.

Example 1-3: Add All Selected Variables and All Values to the DataBasket

This example continues from Step 5 of Example 1-1. Clicking “Browse/Select Highlighted Variables” in that step opens a new window, “Browse/Select Variables & Values.”

Example 1-3: Adding Unrestricted Values		
Step	What to do	Why do it?
1.1 1.2	In the “Browse/Select Variables & Values” window, check the box next to “Select ALL Variables” (1.1) and click “OK” (1.2).	This allows you to add all values for all variables to be added to your DataBasket.
2	A pop-up window will appear that contains information on how many variables you have added to your DataBasket. If the displayed information is correct, click “OK.”	This step gives you the opportunity to confirm the number of variables that is placed in your DataBasket.
3	If the number of variables displayed in the confirmation pop-up window is incorrect, click “Cancel.”	Upon choosing “Cancel,” no variables are placed in your DataBasket and the selections of variables and values will need to be remade.
4.1 4.2	You will be returned to the “Step 1: Select Dataset & Variable” window upon clicking “OK” (Step 2 above). Click the “Step 2: DataBasket/Download/Make a Table” tab (4.1) to see a listing of variables that are in your DataBasket (4.2).	This completes adding unrestricted values for all selected variables to your DataBasket.

Step 1

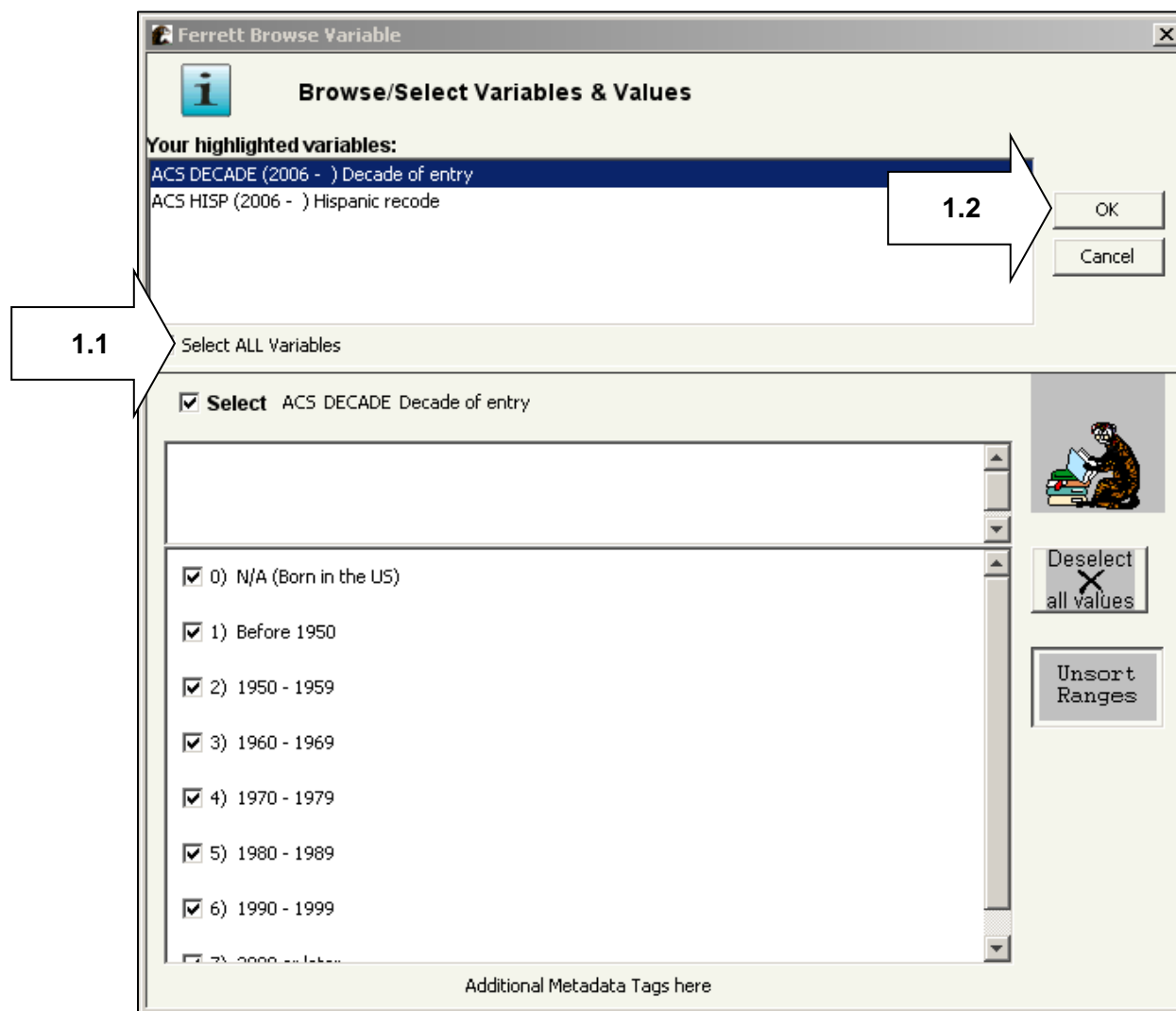


Figure 1-20: Variables Selected for DataBasket

Step 2/3

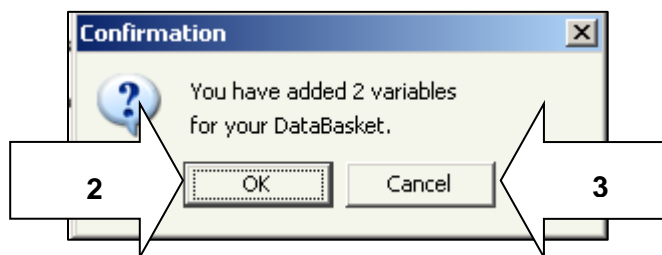


Figure 1-21: Confirmation of Number of Variables in DataBasket

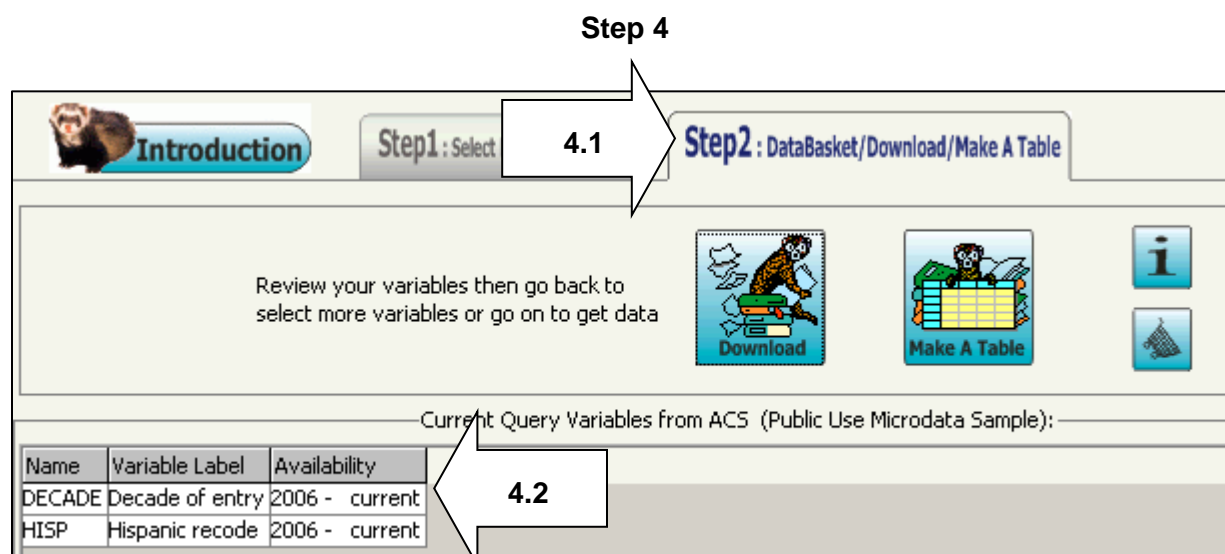


Figure 1-22: Variables in DataBasket

Example 1-4: Add Restricted Values for Individual Variable

This example continues from Step 5 of Example 1-1. Clicking “Browse/Select Highlighted Variables” in that step opens a new window, “Browse/Select Variables & Values.” In this example, we will be restricting the values for the variable “Decade of Entry” to decades before 1980.

Example 1-4: Adding Restricted Values for Individual Variable		
Step	What to do	Why do it?
1.1 1.2	Highlight “Decade of Entry” in the “Browse/Select Variables & Values” window (if not automatically highlighted) (1.1) and note that the box next to “Select ACS DECADE Decade of Entry” has been automatically checked (1.2).	DataFerrett defaults to automatically selecting the first variable on the list of selected variables (Figure 1-7).
2	Click on “Deselect All Values”.	In order to restrict values associated with individual variables, you must first clear all the default selected variables.
3.1 3.2	Click on boxes next to values 0 – 4 (3.1) and click “OK” (3.2).	This step specifies values to be placed in the DataBasket.
4	A pop-up window will appear that contains information on how many variables you have added to your DataBasket. If the displayed information is correct, click “OK.”	This step gives you the opportunity to confirm the number of variables you wish to be placed in your DataBasket.
5	If the number of variables displayed in the confirmation pop-up window is incorrect, click “Cancel.”	Upon choosing “Cancel,” no variables are placed in your DataBasket and the selections of variables and values will need to be remade.
6.1 6.2	You will be returned to the “Step 1: Select Dataset & Variable” window upon clicking	This confirms the variables that are currently in your DataBasket.

	"OK" (Step 2 above). Click the "Step 2: DataBasket/Download/Make a Table" tab (6.1) to see a listing of variables that are in your DataBasket (6.2).	
7.1 7.2	Highlight the "Decade of Entry" variable in the "Current Query Variables" section of the "Step 2" tab (7.1) and click "View/Modify Variable(s)" in the "Act on Your Query" listing (7.2).	This step allows you to see the restricted values for an individual variable that is currently in your DataBasket.

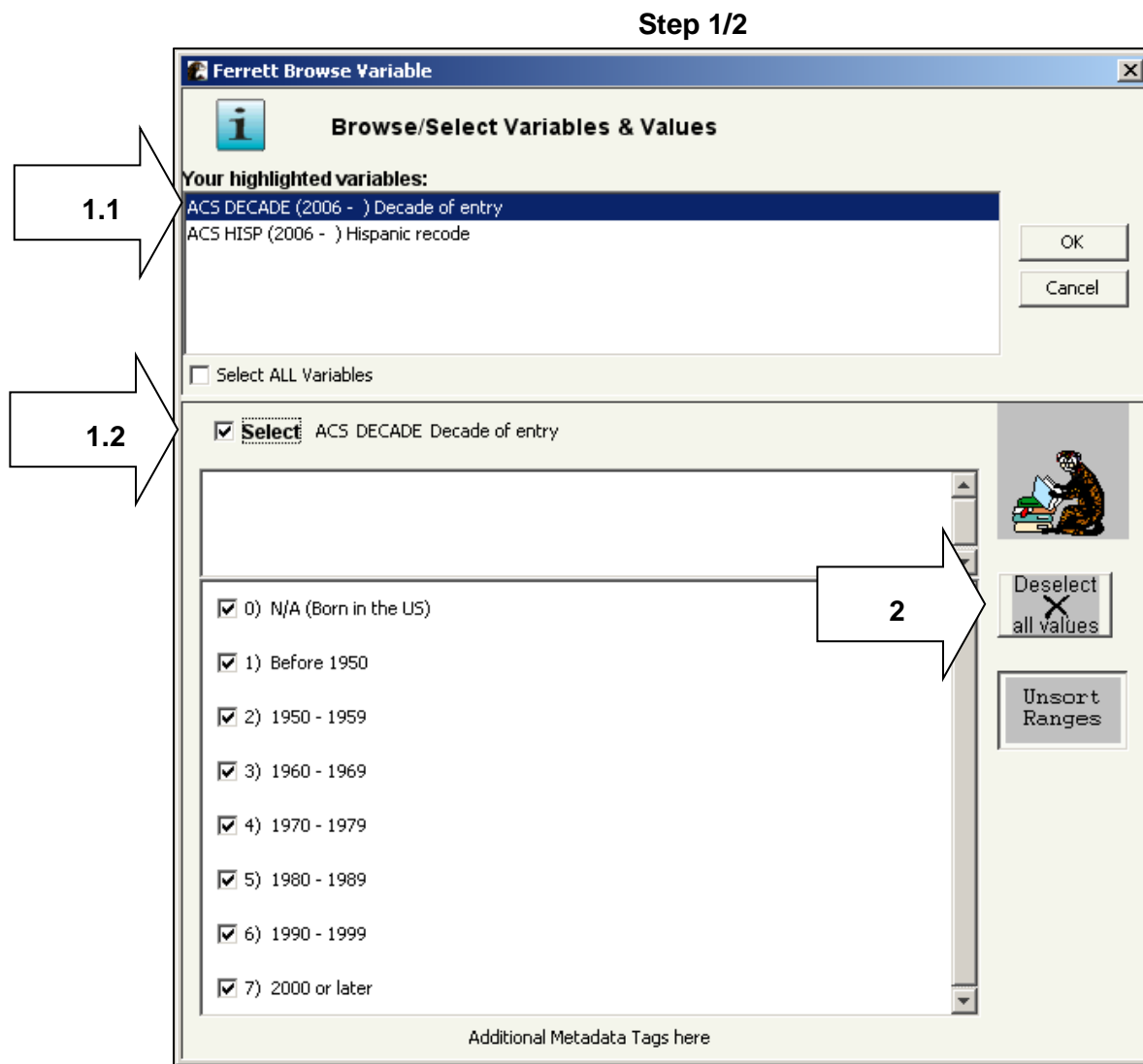


Figure 1-23: Variable Selected for DataBasket

Step 3

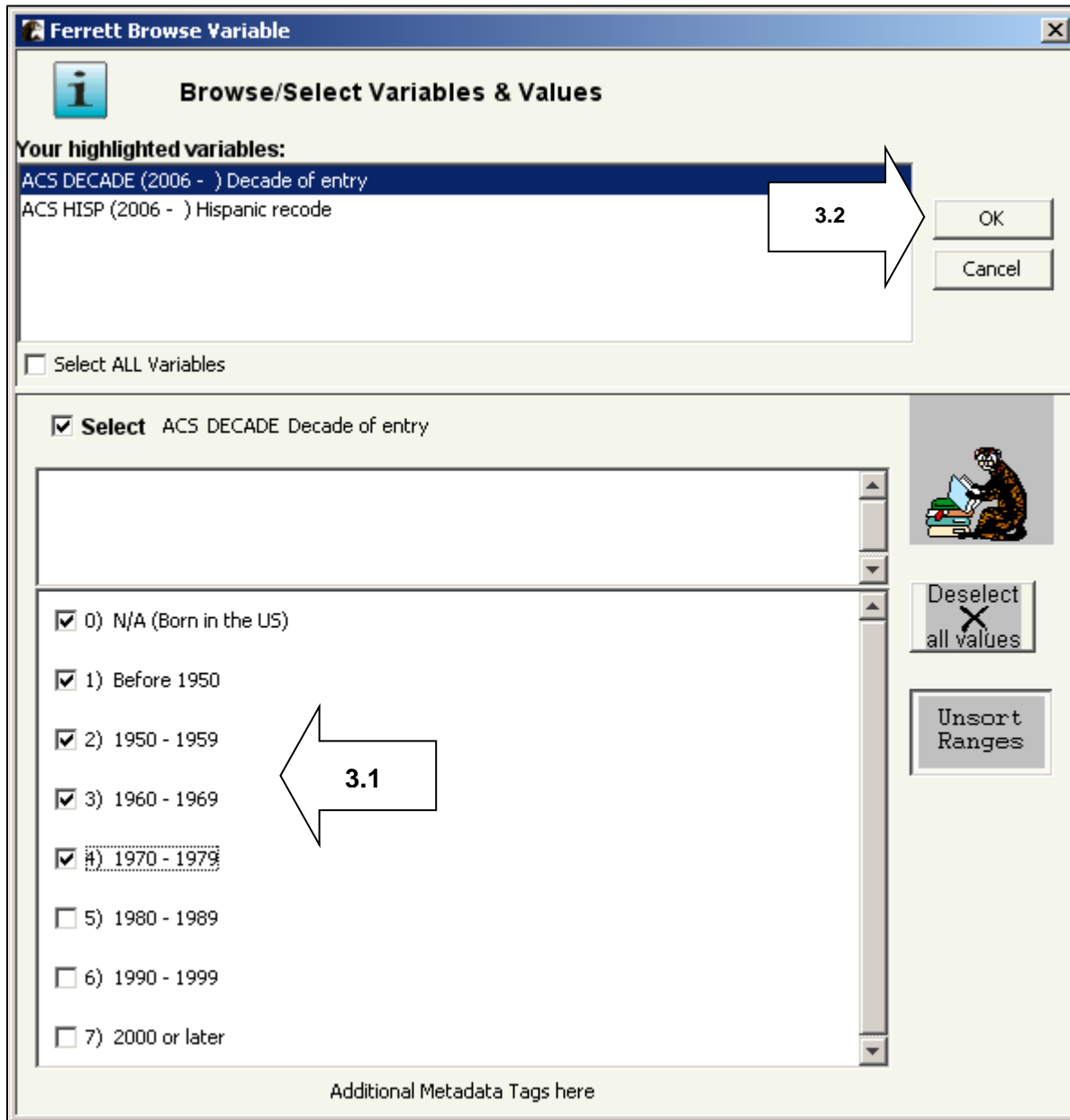


Figure 1-24: Variable Values Restricted

Step 4/5

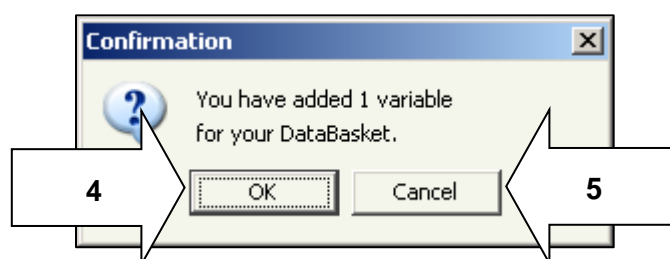


Figure 1-25: Confirmation of Number of Variables in DataBasket

Step 6

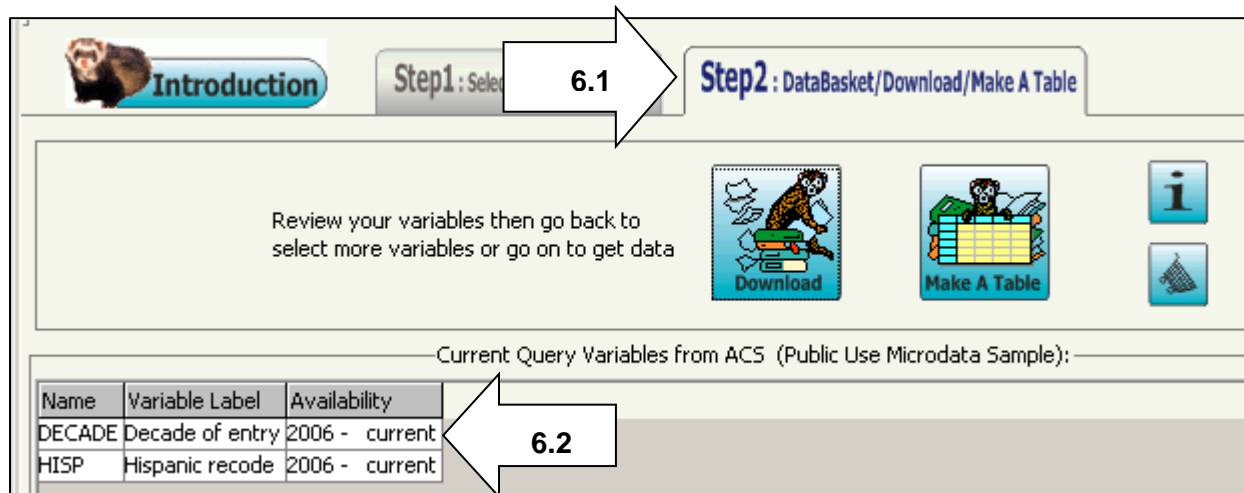


Figure 1-26: Variables in DataBasket

Step 7

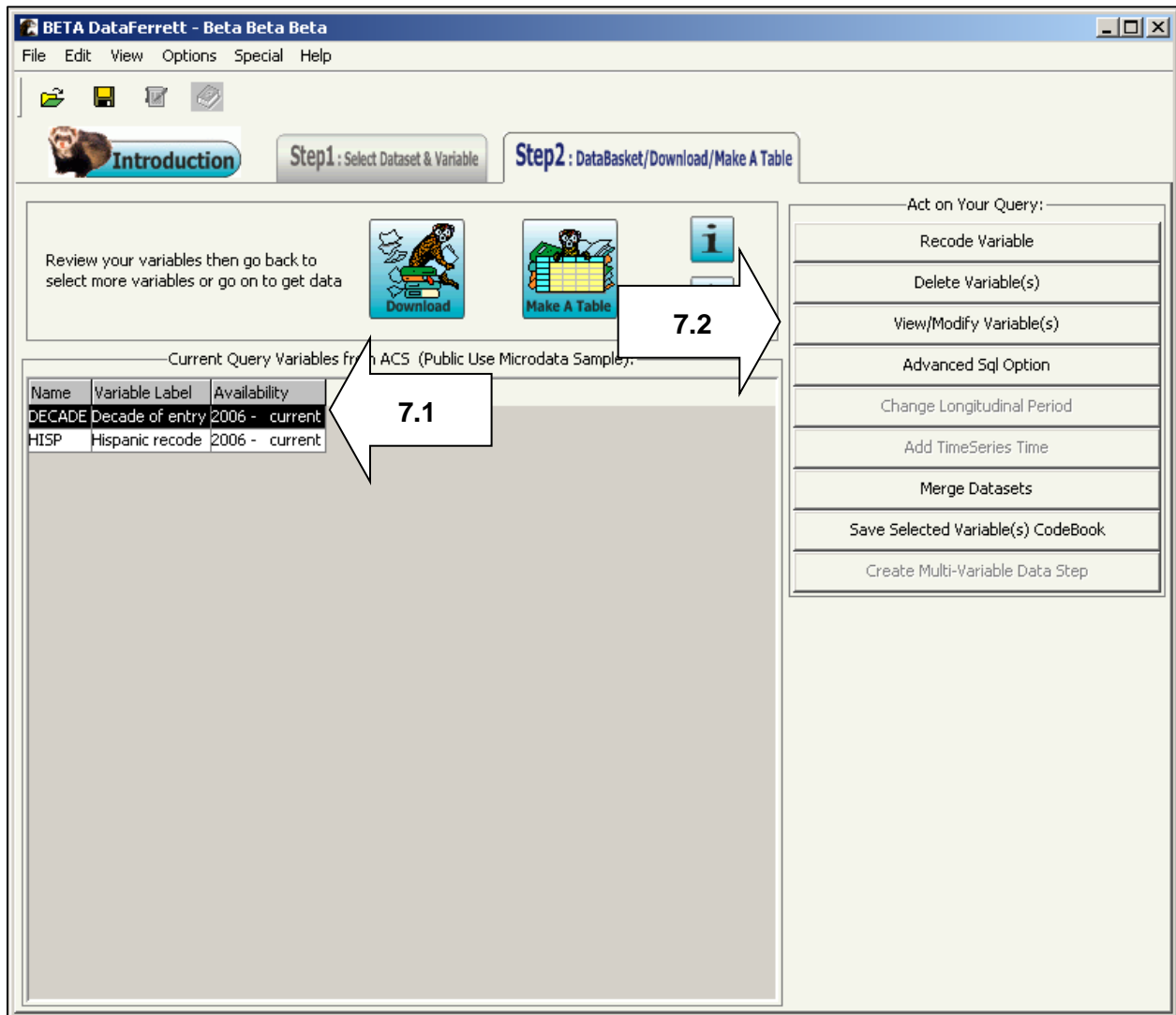


Figure 1-27: View Restricted Values for Variable

Example 1-5: Add Restricted Values for All Variables to the DataBasket

This example continues from Step 6 of Example 1-3. At the conclusion of this example, values will be restricted for the “Decade of Entry” variable to decades before 1980 and to South American countries for the “Hispanic Recode” variable.

Example 1-5: Adding Restricted Values for All Variables		
Step	What to do	Why do it?
1.1 1.2	Highlight the “Hispanic Recode” variable in the “Current Query Variables” section of the “Step 2” tab (1.1) and click “View/Modify Variable(s)” in the “Act on Your Query” listing (1.2).	This step opens the “Browse/Select Variables & Values” window that shows the values associated with the individual variable. Note that all values for the “Hispanic Recode” variable are checked.
2	Click on “Deselect All Values”.	In order to restrict values associated with individual variables, you must first eliminate ALL variables from being selected.
3.1 3.2	Click on boxes next to values 13 - 22 (3.1) and click “OK” (3.2).	This step specifies values to be placed in the DataBasket.
4	A pop-up window will appear that contains information on how many variables you have added to your DataBasket. If the displayed information is correct, click “OK.”	This step gives you the opportunity to confirm the number of variables you wish to be placed in your DataBasket.
5	If the number of variables displayed in the confirmation pop-up window is incorrect, click “Cancel.”	Upon choosing “Cancel,” no variables are placed in your DataBasket and the selections of variables and values will need to be remade.
6.1 6.2	You will be returned to the “Step 2: DataBasket/Download/Make a Table” tab (6.1) to see a listing of variables that are in your DataBasket (6.2).	This confirms the variables that are currently in your DataBasket.
7.1 7.2	Highlight the “Decade of Entry” variable in the “Current Query Variables” section of the “Step 2” tab (7.1) and click “View/Modify Variable(s)” in the “Act on Your Query” listing (7.2). Repeat these steps for the “Hispanic Recode” variable.	This step allows you to see the restricted values for all variables that are currently in your DataBasket.

Step 1

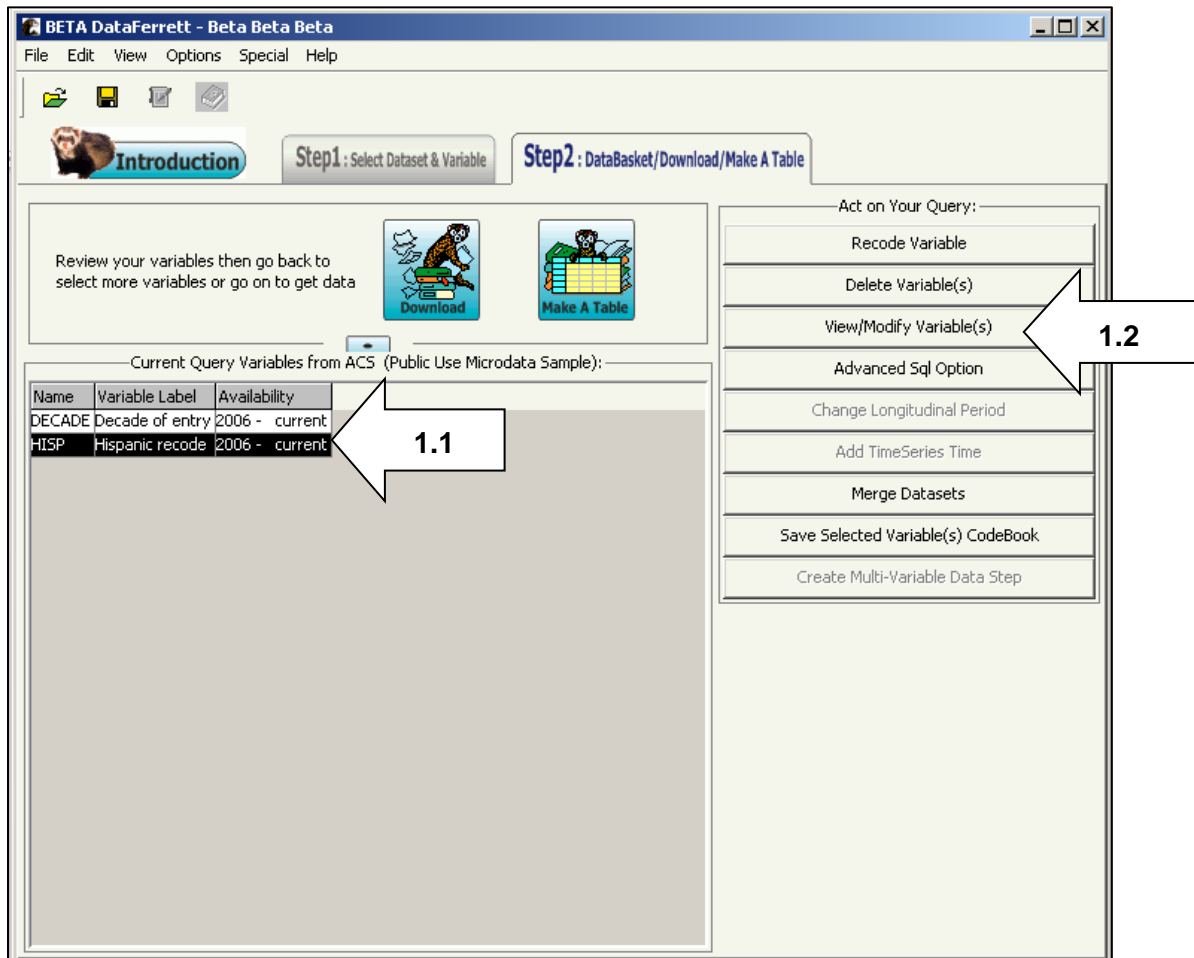


Figure 1-28: View Values for Variable

Step 2

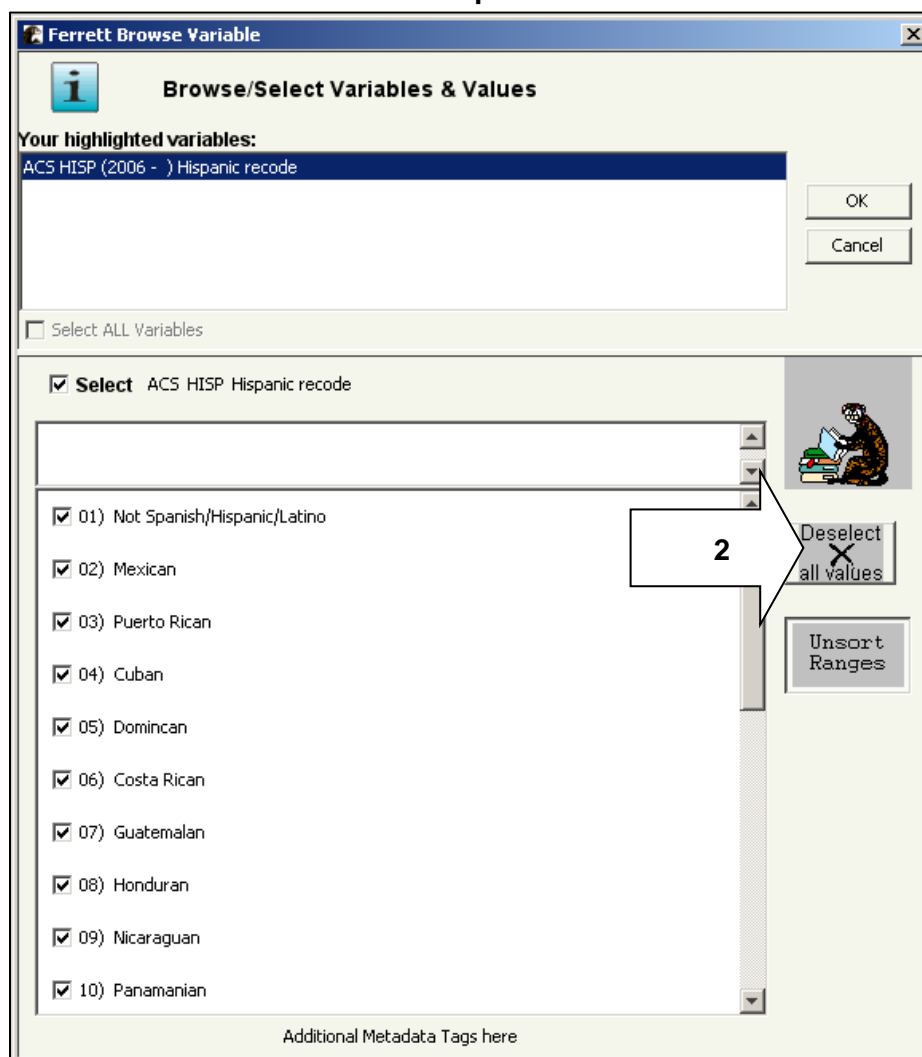


Figure 1-29: First Step in Restricting Values

Step 3

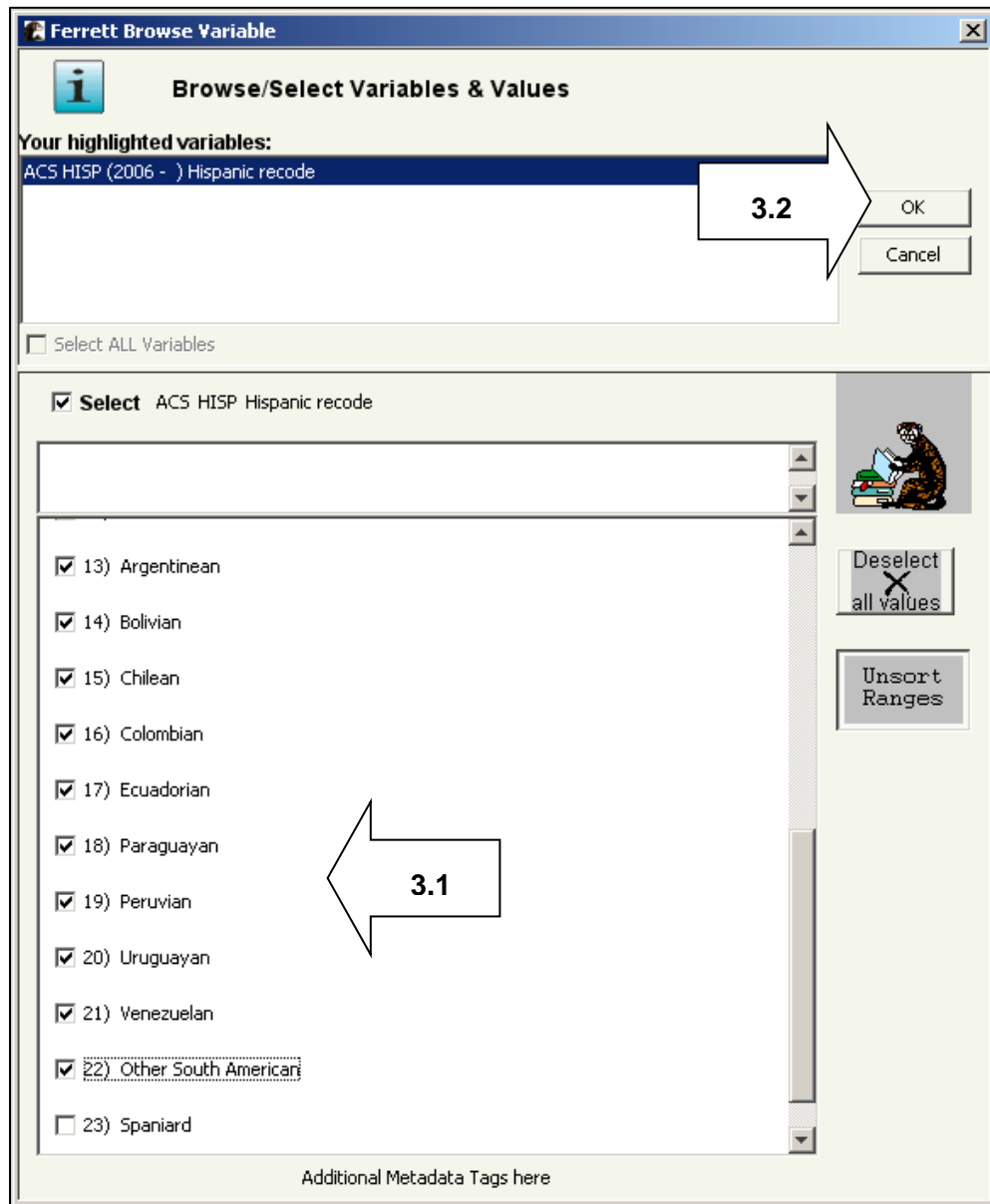


Figure 1-30: Second Step in Restricting Values

Step 4/5

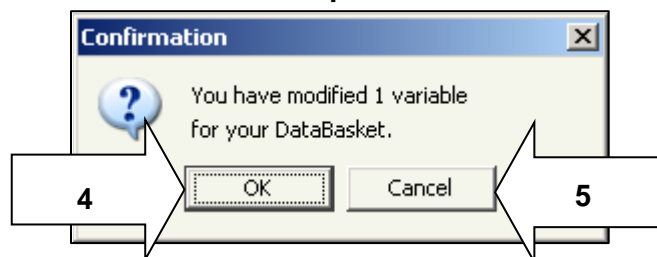


Figure 1-31: Confirmation of Number of Modified Variables in DataBasket

Step 6

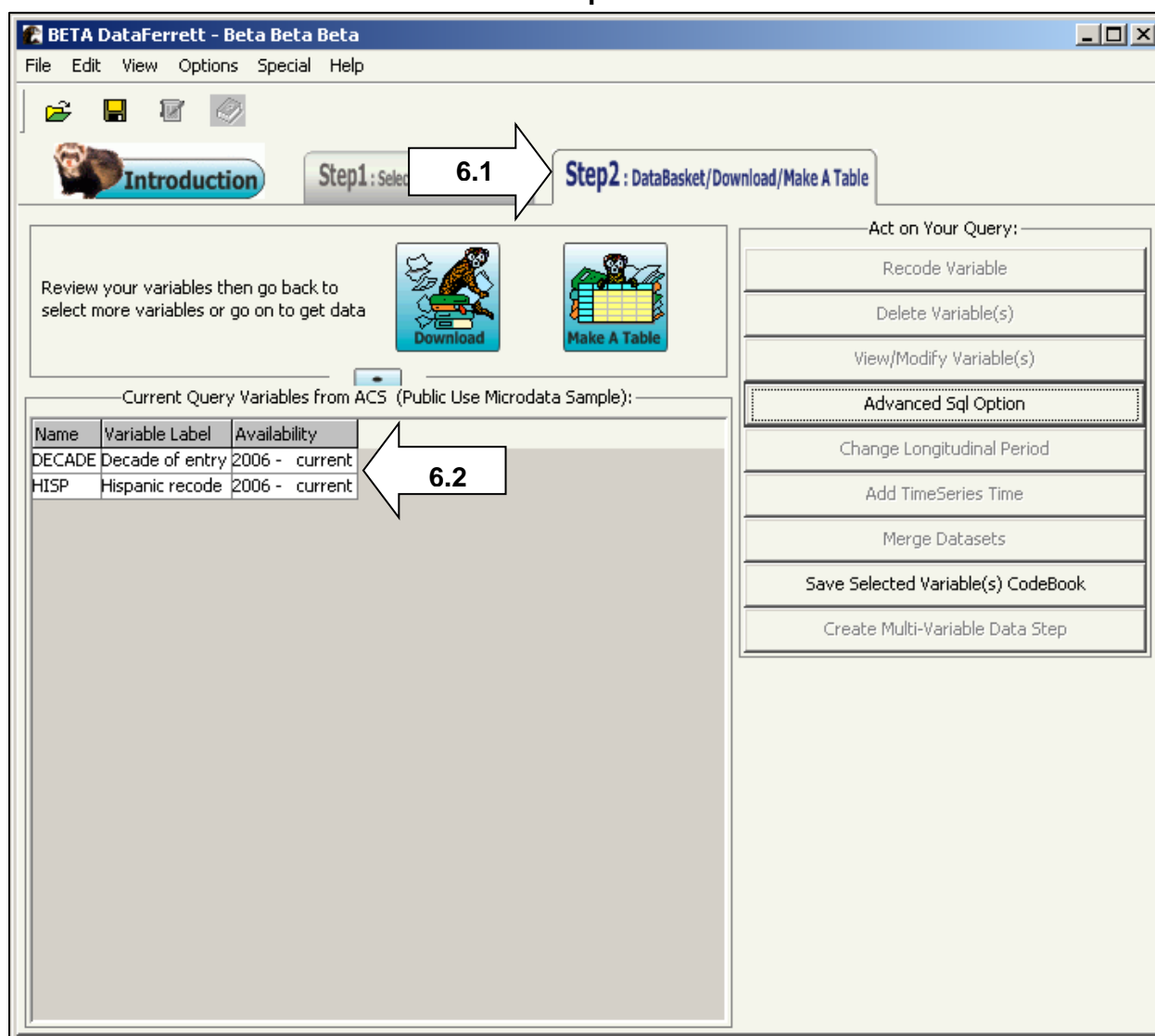


Figure 1-32: Variables in DataBasket

Step 7

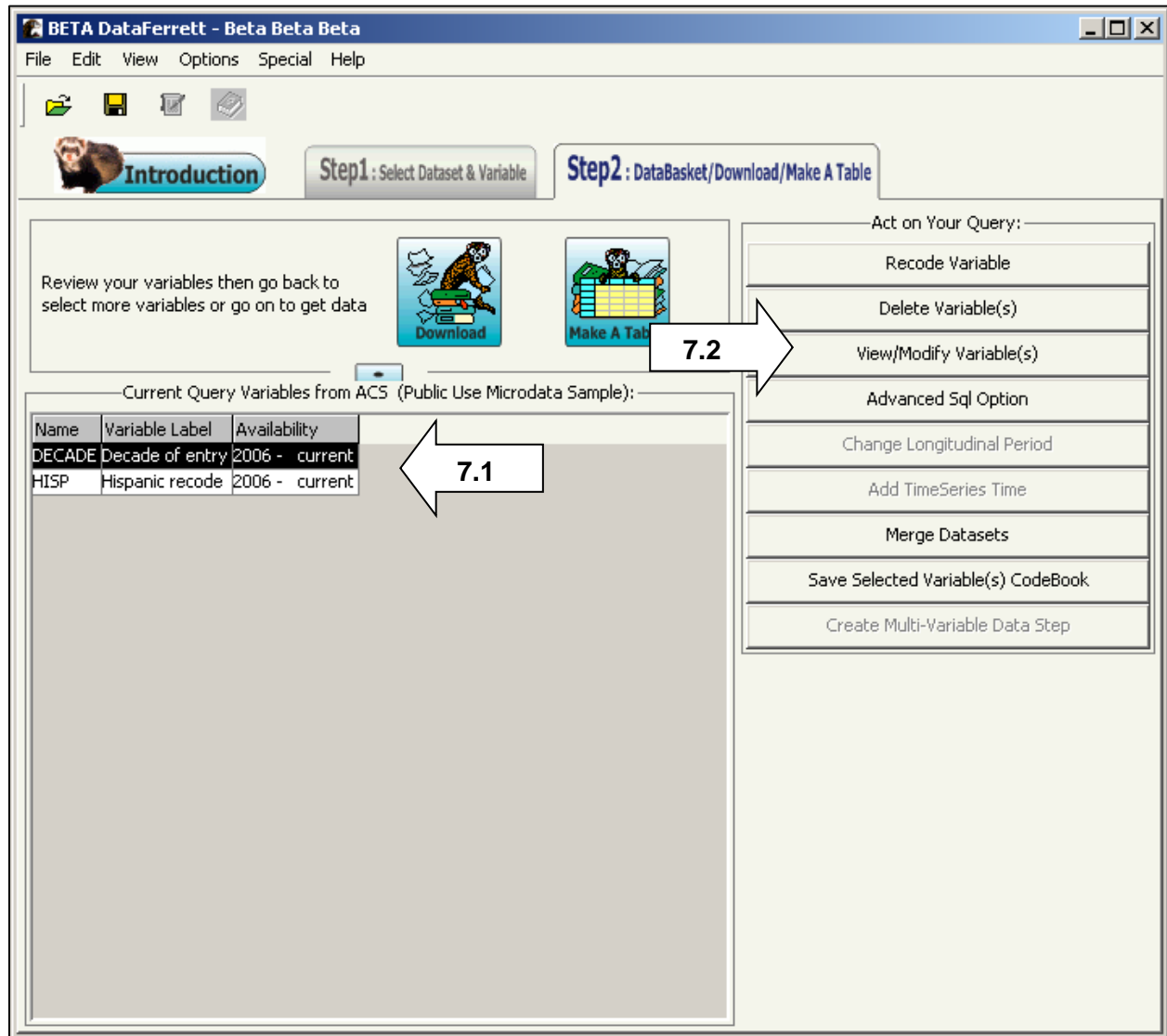


Figure 1-33: View Restricted Values for Variable

"Step 1" – Other Options

While the main purpose of the "Step 1: Select Dataset & Variables" window is to search for variables within datasets and add them to the DataBasket, other helpful options (described in Figure 1-34 below) are also available to the user.

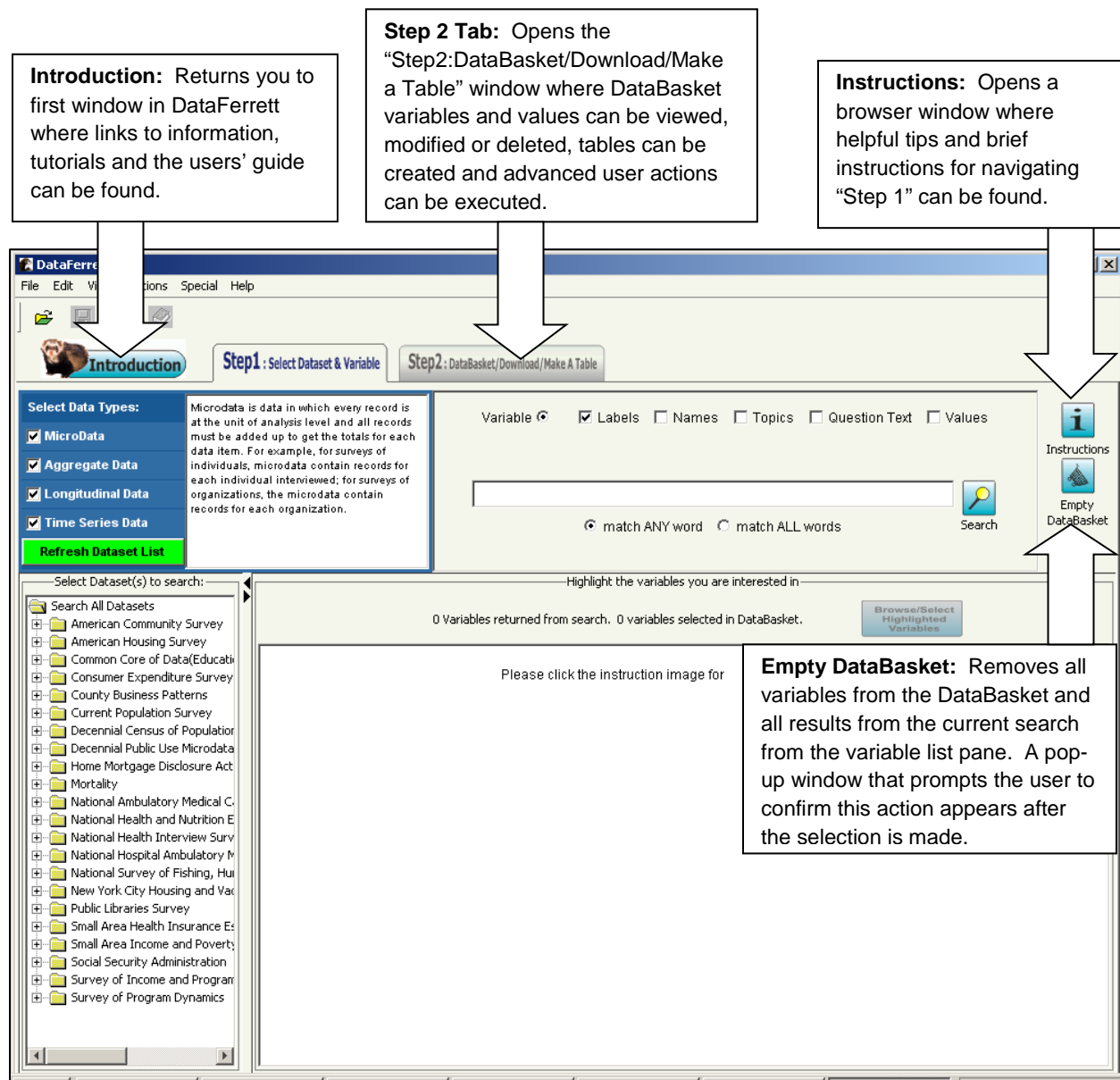


Figure 1-34: Additional User Options

Menu Options

Menu options, located at the top left of the "Step 1" screen, provide expanded choices to the user while working in DataFerrett. Each of these options is briefly described below.

File Menu



Open: Opens a saved/existing Ferrett Session File (fsf) or a Ferrett Tabulation File (ftf)

Save As: Saves Ferrett Session Files

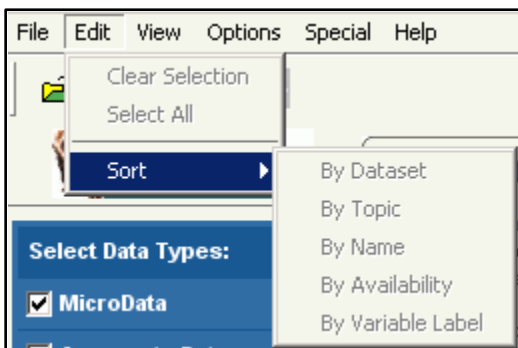
Rename: Not available at this time

Debug: Available only to authorized users

Print Debug List: Not available at this time

Exit: Closes DataFerrett with option of saving current session

Edit Menu



Clear Selection: Not available at this time

Select All: Highlights all variables in the selection pane

Sort: Provides options for sorting variables:

By Dataset: Not available at this time

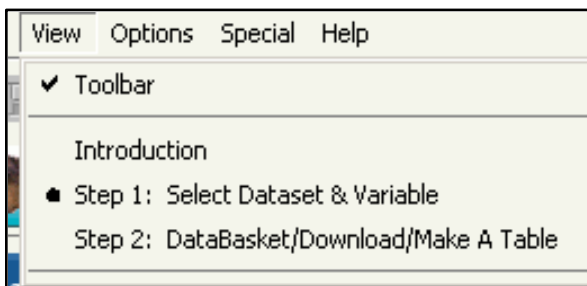
By Topic: Alphabetical listing

By Name: Alphabetical listing

By Availability: Chronological listing

By Variable Level: Alphabetical listing

View Menu



Toolbar: Toggles the toolbar between hidden and seen

Introduction: Opens a browser window for helpful tips and instructions

Step 1: Opens the “Step 1” window; dot indicates window is active

Step 2: Opens the “Step 2” window; dot indicates window is active

Options Menu



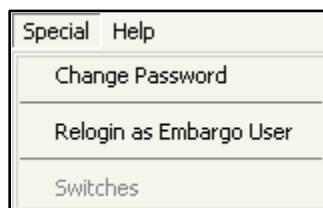
Item Types: Narrows or expands the number of records from a variable search. If an item type is changed, variables will need to be reselected.

General Items: Deselecting this option reduces the number of records returned from a search

Flags: Applicable to select datasets

Weights: Represents a sample from various populations

Special Menu



Change Password: Generates a blank text box in which a new password can be entered

Relogin as Embargo User: Available only to those users with special permission to view pre-released data

Switches: Not available at this time

Help Menu



Contents and Index: Not available at this time

How Do I?: Not available at this time

Glossary: Not available at this time

Instructions: Not available at this time

Tell Me About: Not available at this time

Ferrett Tabulation Examples: Provides examples of pre-composed statistical tables and how each one is created and populated

Ferrett Quick Tour: Basic introduction to DataFerrett

Help Request: Opens browser window to submit requests and/or error reports

Known Bugs: Opens browser window that explains known bugs and workarounds

Release Notes: Opens browser window that reports new features and error corrections executed in latest release

Users Guide: Opens browser window that features the DataFerrett Users Guide

About Ferrett: Opens browser window that provides brief description of DataFerrett

CHAPTER 2: BROWSE/SELECT GEOGRAPHY VARIABLES

As explained in the previous chapter, variables are characteristics found within each dataset. While several datasets may contain common variables, certain variables such as geography are unique to select datasets. In addition, due to the nature of geographic definitions, the approach taken to adding geographic variables to the DataBasket is quite different from the approach taken for other types of variables. The “Ferrett Geography Codebook” defines this approach, which is a streamlined process of selecting geographic areas to add to the DataBasket.

Introduction to Ferrett Geography Codebook Interface

This chapter describes the geography selection function in DataFerrett, which can be used to organize data in a dataset by type of geography and/or to restrict the data to certain geographic areas. This chapter will address the following:

- The geography selection window, which may be new to experienced DataFerrett users;
- The selection of geographic variables based on type of dataset being used;
- Basic functionality of the geographic selection process; and
- Examples that demonstrate different scenarios for selecting geographic variables.

The Geography Selection Window

The geography selection window, referred to as the “Ferrett Geography Codebook,” is divided into three panes and is designed to use both “drag and drop” or “double click” functionality. The left pane lists all of the geographies available for the selected dataset. The middle pane allows a user to drill down to more specific geographic levels, referred to as “hierarchies”, that may be a part of a selected geography. The right pane is where the final geographic selection is made and added to the DataBasket. The Browse/Select Geographies window for the American Community Survey (ACS) 5-Year Summary File is shown in Figure 2-1 below.

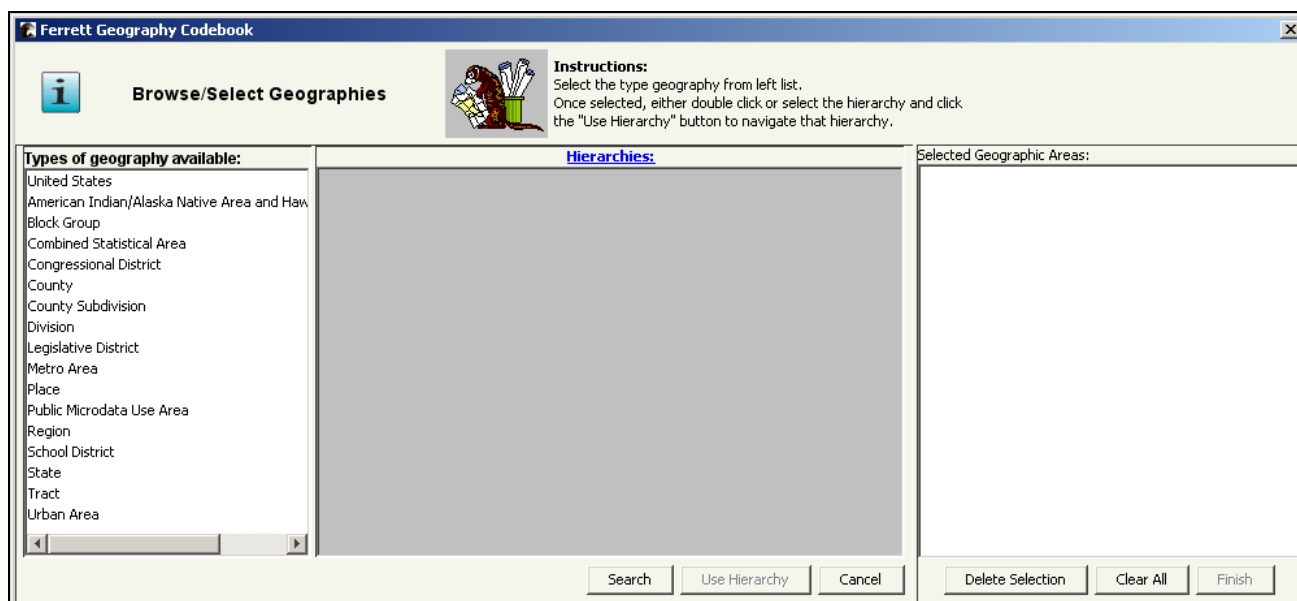


Figure 2-1: ACS 5-year Summary File List of Available Geographies shown in the “Ferrett Geography Codebook”

NOTE: Different geographies are available depending on the dataset in use. After selecting the dataset, the geographies that are available will appear in the left-hand pane of the geographic interface.

Selection of Geographic Variable

Some datasets in DataFerrett *require* a geographic variable to be selected in order to proceed to “Step 2: DataBasket/Download/Make a Table”, while others do not. Most aggregate, or pre-summed, datasets require a geographic selection. DataFerrett will notify the user that a geographic variable must be selected for these datasets before proceeding to Step 2. Other datasets do not require a geographic variable to be selected, such as Public Use Microdata Sample (PUMS); however, there may be different levels of geography associated with the dataset that can be selected as an *option* for lower-level geographic analysis. The process for selecting geographies within the geographic selection interface for either type of dataset remains the same, but the process for *accessing* the geographic selection interface for each type of dataset differs.

For datasets that *require* the selection of a geographic variable, such the ACS 5-Year Summary File, DataFerrett alerts the user that geographic variables must be selected (Figure 2-2). The user must click “OK” in response to this alert before proceeding to the “Required Variables” window shown in Figure 2-3.



Figure 2-2: Geographic Variable Alert

Highlight the variables you are interested in

Currently viewing: B01001A. SEX BY AGE (WHITE ALONE) 31 Variables returned from search. 2 variables selected in DataBasket.

Browse/Select Highlighted

REQUIRED VARIABLES					
Selected	Topic	Name	Availability	Variable Label	
<input checked="" type="checkbox"/>	Default values selected for you	Selectable Geographies COMPONENT	2005-2009 - current	Geographic Component	
<input checked="" type="checkbox"/>	Needs to be selected	Selectable Geographies Geography	2005-2009 - current	Geographic Items	

Figure 2-3: Required Variables Window

After double clicking on the “Geography” variable name in the “Needs to be selected” row, the geographic selection user interface (referred to as the “Ferrett Geography Codebook,” Figure 2-1) will open.

For datasets that have a geographic selection as an *option*, the process for accessing the Ferrett Geography Codebook is different. For example, when working with the ACS PUMS Sample 1 year data, the geography variable appears within the list of selected variables, as opposed to a separate “Required Variables” window as demonstrated above. Figure 2-4 below illustrates the geographic variable within the list of all variables a user has selected for browsing. Once the list of variables appears after clicking on “Search Variables,” you can either highlight the Geographic Items variable label and click on the “Browse/Select Highlighted Variables” button on the upper right, or double-click on the variable itself – either option will open the Ferrett Geography Codebook.

NOTE: Make sure the “Selectable Geographies” topic is checked along with any other topics in which you are interested prior to clicking “Search Variables.”

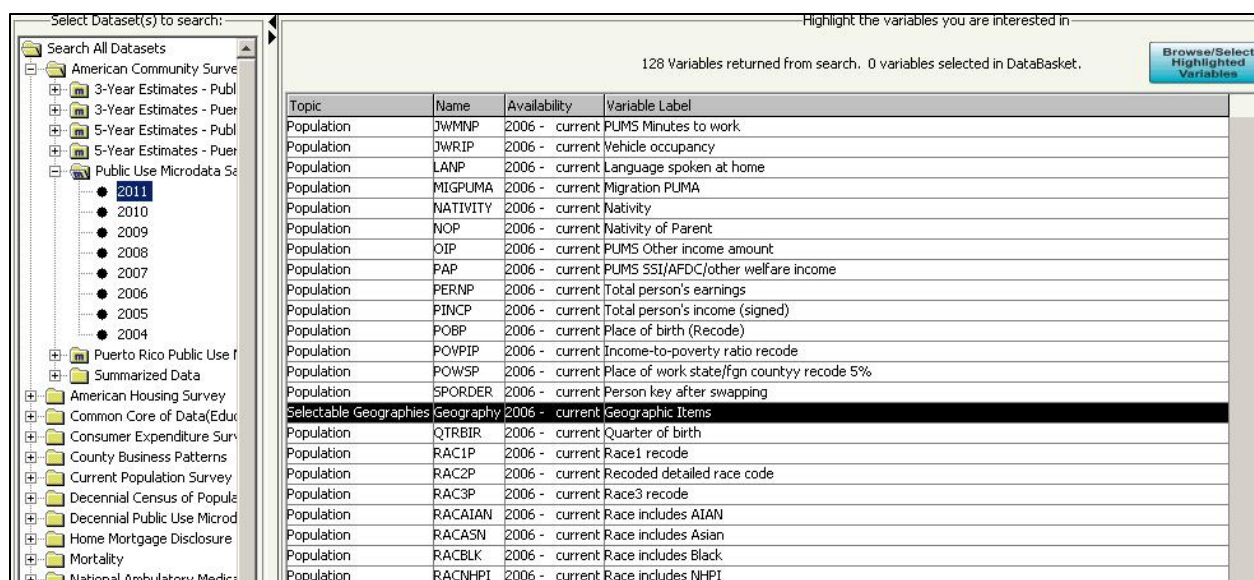


Figure 2-4: Geography/Geographic Items Variable Highlighted within the Variable Selection List.

Geographic Selection Process Functionality

Once the Ferrett Geography Codebook is open, the geographic selection process is determined by the type of geography available for a selected dataset and the geographic hierarchies, or levels of geography, associated with the dataset. (Official Census Bureau hierarchies are also known as “geographic summary levels.”) For example, counties are contained within the State-County hierarchy. Geography selection begins by clicking on a type of geography in the left pane and choosing one of the hierarchies associated with that geography type from the list that appears in the middle pane. The layout gives you a preview of the geographic data that is available in each area by highlighting it. Figure 2-5 below depicts the geographic data available associated with the County Business Patterns county level dataset.

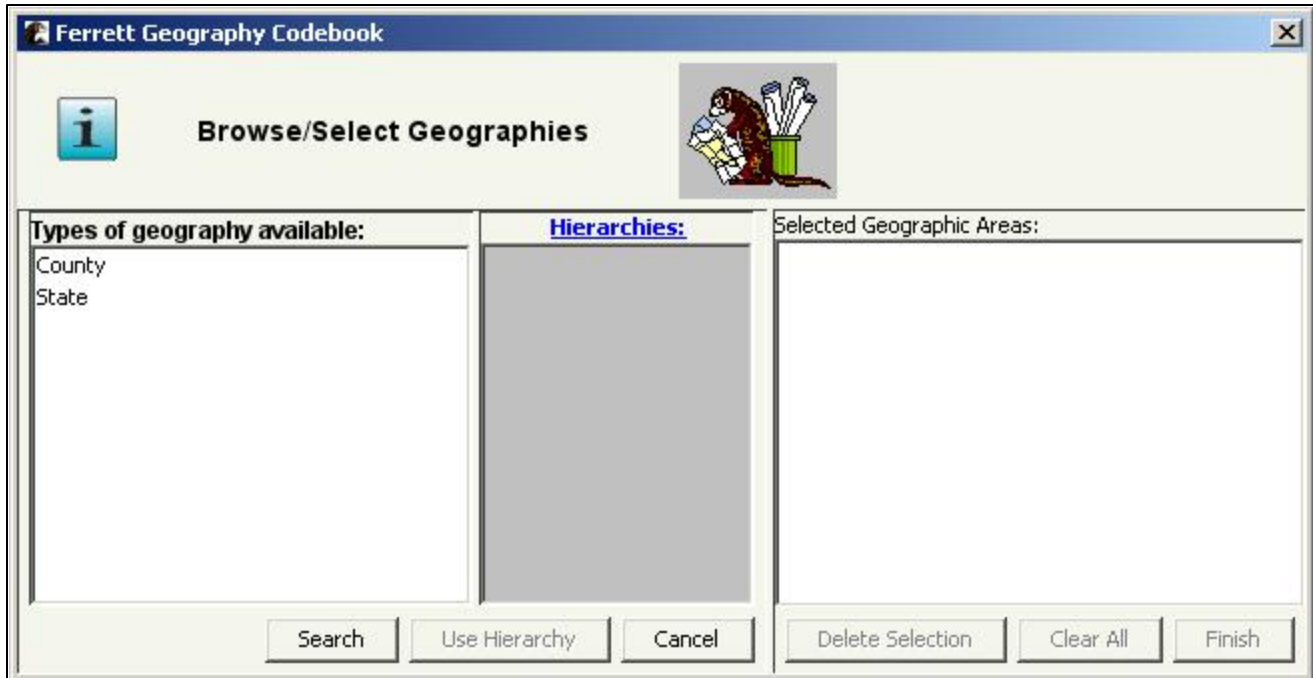


Figure 2-5: Types of Geography Available for County Business Patterns County-Level Dataset

After highlighting a type of geography from the list in the left pane, the hierarchies associated with it are shown in the middle pane making it easier to choose the correct geographic area. As shown in Figure 2-6 below, the hierarchies associated with the State type of geography within the County Business Patterns dataset are FIPS State Code and FIPS State Code > FIPS County Code.



Figure 2-6: Types of Hierarchies Available for County Business Patterns State Geography Selection

The next step in selecting a geographic area is to highlight the desired hierarchy from the middle pane and click “Use Hierarchy.” This action generates a new window in which a list of specific geographies is displayed in the left pane. In Figure 2-7 below, FIPS State Code > FIPS County Code is chosen as the hierarchy, which opens a new window in which the specific geography of Alabama is highlighted from the list in the left pane (Figure 2-8).

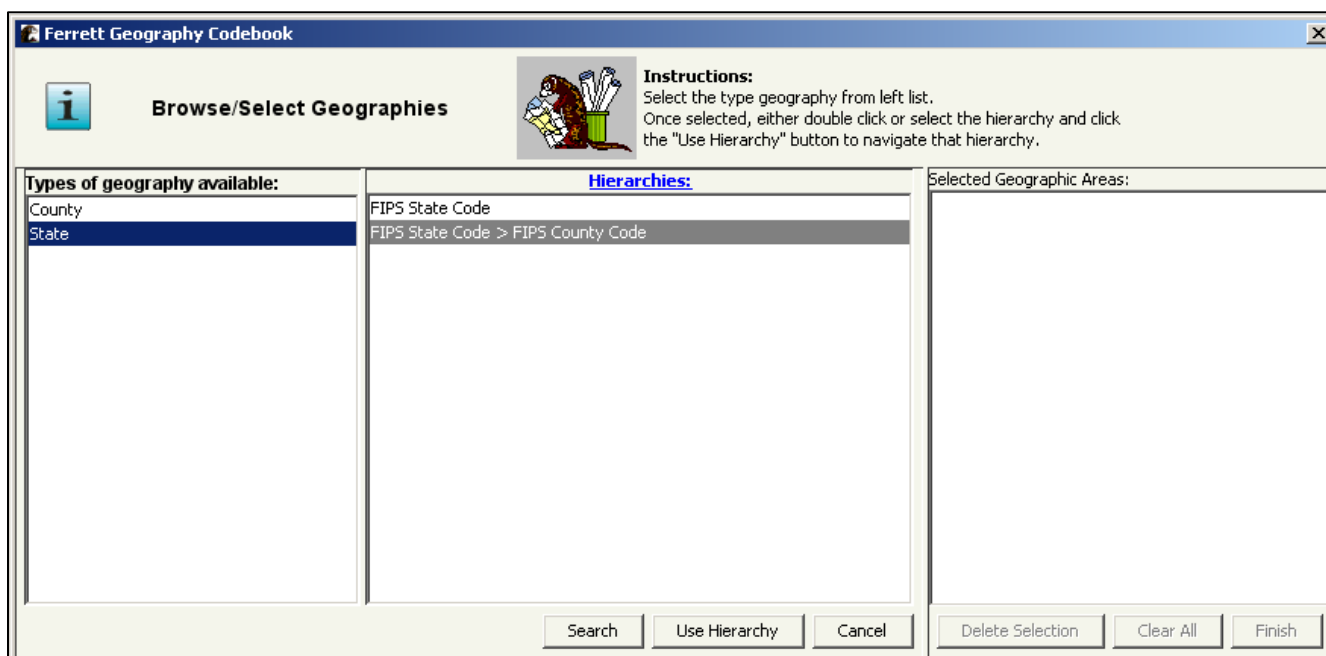


Figure 2-7: Highlight “FIPS State Code>FIPS County Code” and then click “Use Hierarchy”

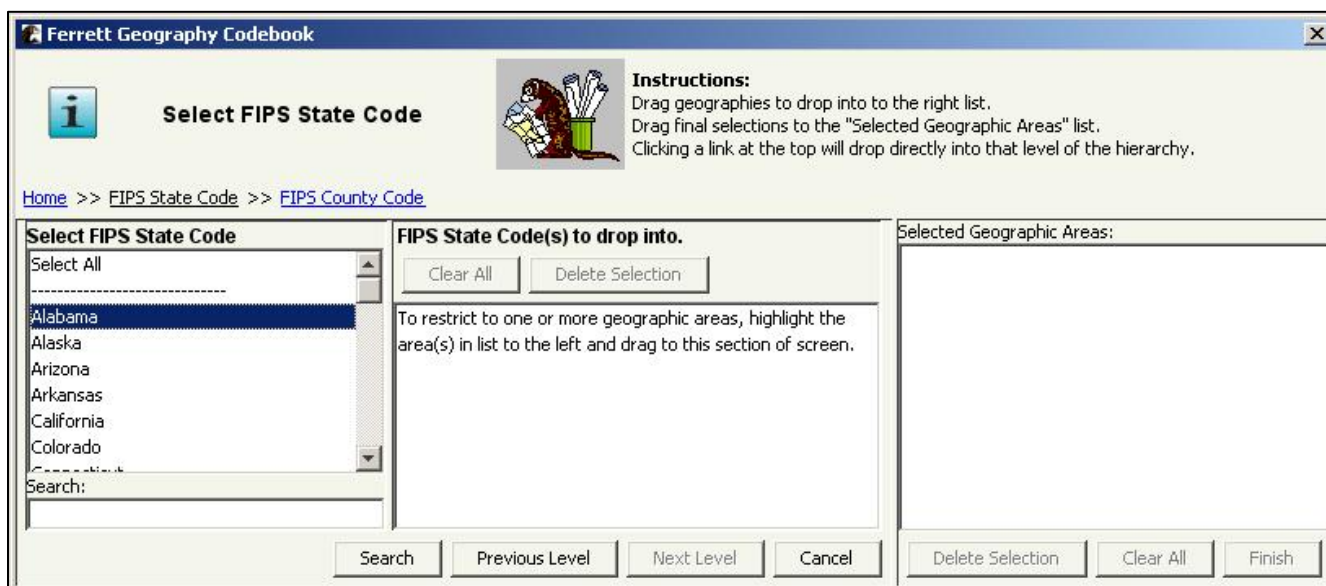


Figure 2-8: Selecting Alabama for County Business Patterns County Level Dataset

At this point, the highlighted state of Alabama can be dragged directly into the “Selected Geographic Areas:” pane on the right side of the window. Doing so will return the state level geography for inclusion in the DataBasket (Figure 2-9).

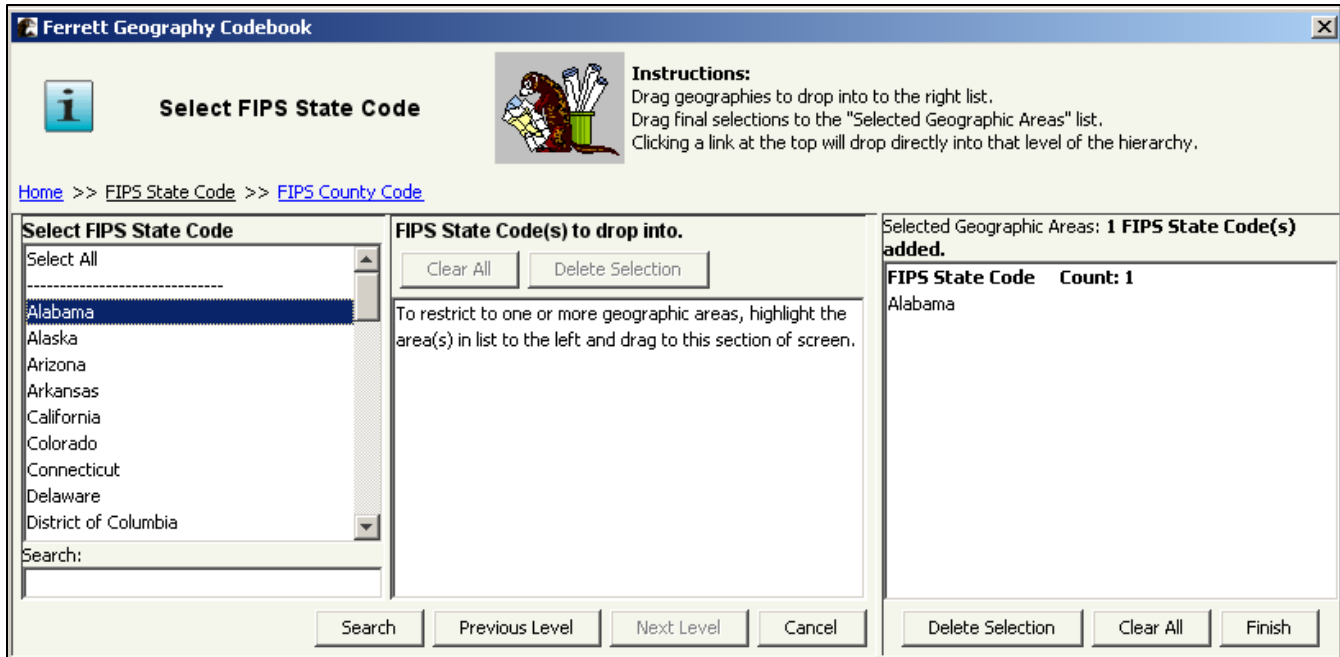


Figure 2-9: Selecting State Level Geography for DataBasket

Conversely, all of the counties for Alabama (the county level hierarchy associated with the state geography) can be selected by highlighting the state of Alabama and either double clicking it or dragging it into the middle pane “FIPS State Code(s) to drop into”. The next step is to highlight the Alabama selection in the middle “FIPS State Code(s) to drop into” pane and click on “Next Level” as shown in Figure 2-10 below. Doing so will display all of the counties for Alabama in the left “Select FIPS County Code” pane of the Ferrett Geography Codebook. (Figure 2-11).



Figure 2-10: Selecting Alabama for County Business Patterns County Level Dataset

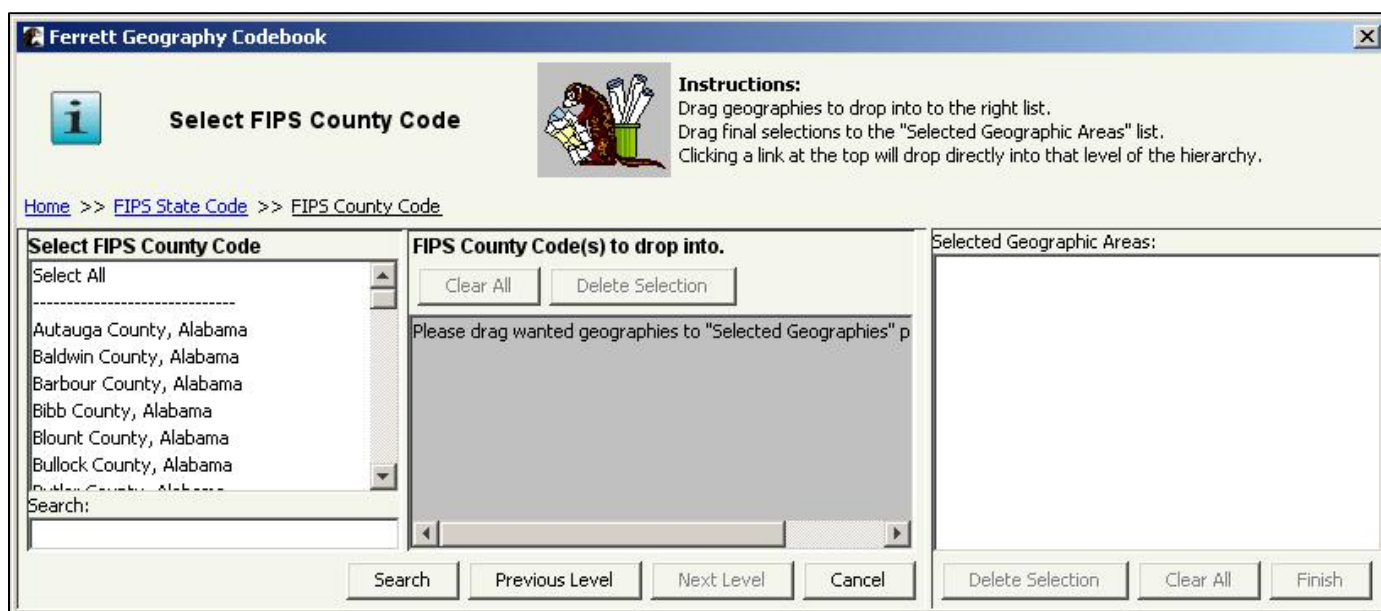


Figure 2-11: After Selecting Alabama for County Business Patterns County Level Dataset

To select the counties in Alabama for inclusion in the DataBasket, either “Select All” or select individual or groups of counties by highlighting them and dragging them into the right “Selected Geographic Areas:” pane. (Figure 2-12).

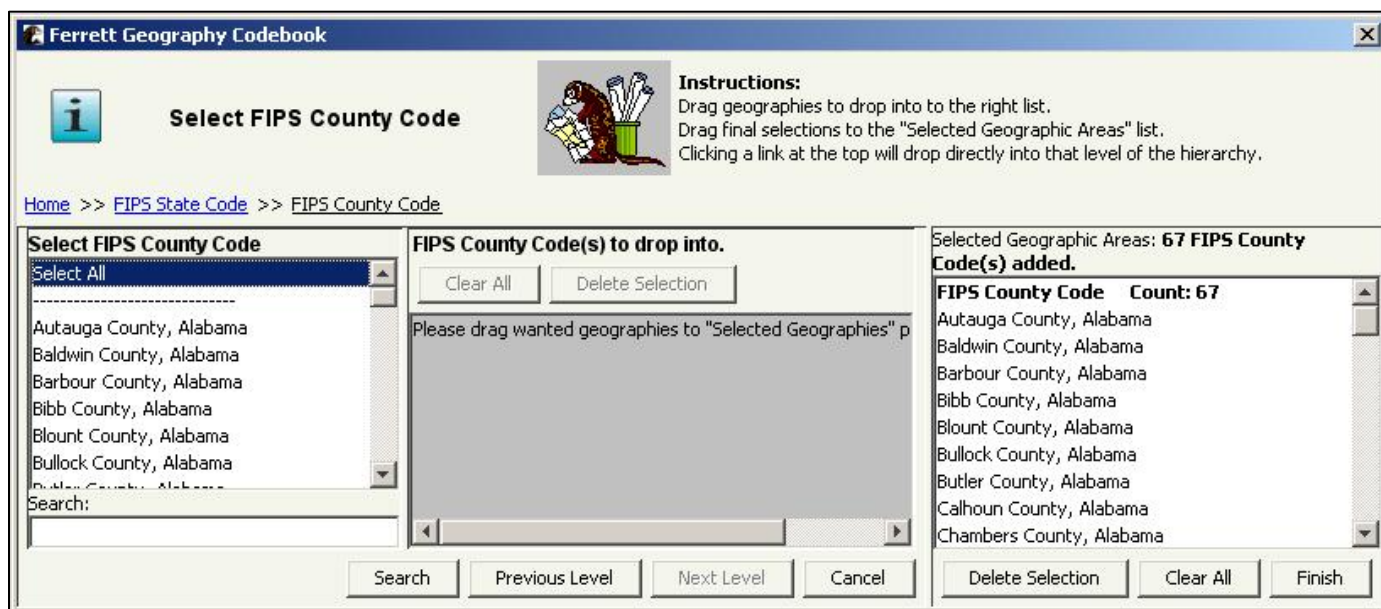


Figure 2-12: Selecting All Counties in Alabama and Dragging Them to Selected Geographic Areas Pane

All of the counties in Alabama can now be added to the DataBasket by clicking “Finish.” If it is necessary to limit the selection of counties, highlight those that are not needed and click on the “Delete Selection” button. To start over in terms of “Selected Geographic Areas” and return to re-select from the left pane, click on the “Clear All” button.

To add additional geographic hierarchies, click on “Previous Level” or use the breadcrumb trail highlighted in Figure 2-13 below BEFORE clicking “Finish”. Doing so allows other geographic hierarchies that are available for that dataset to be added. In this example, the “Previous Level” button was clicked which returned the geographic selection interface to the State hierarchy in the left pane. Alaska was then highlighted and dragged into the “Selected Geographic Areas” pane, which shows that the state of Alaska has been added to the Selected Geographic Areas for inclusion in the DataBasket.

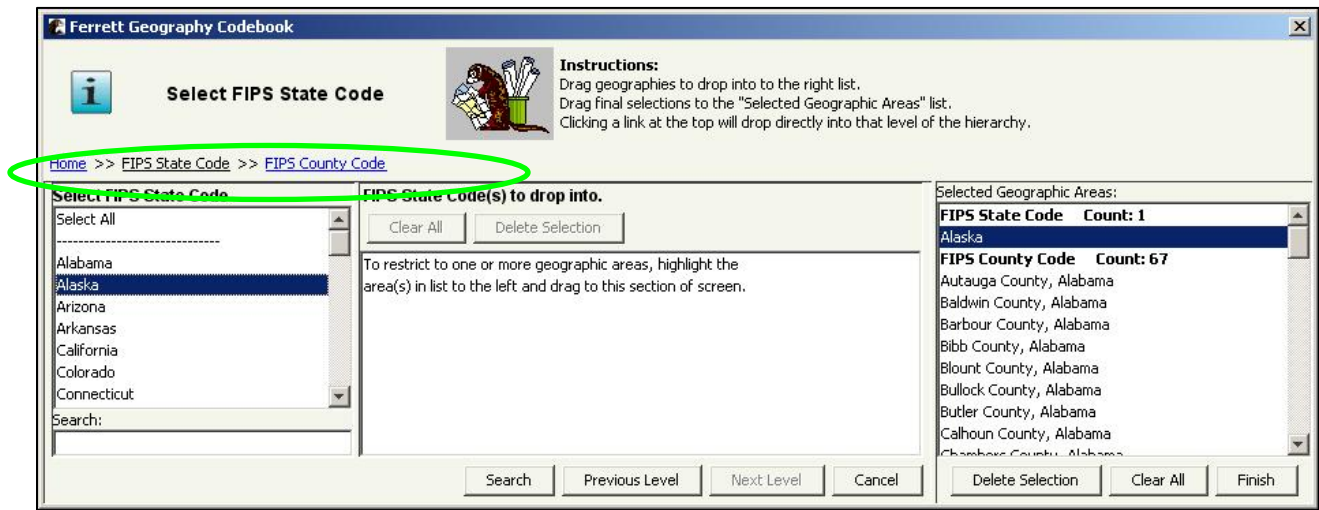


Figure 2-13: Using “Previous Level” to Choose Additional Geography/Hierarchy to Add to Geographic Selection

The following provide step-by-step instructions utilized in adding geographic areas to the DataBasket. Example 2-1 illustrates adding a single geographic area; Example 2-2 demonstrates adding multiple geographic areas.

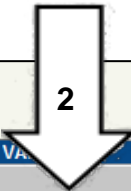
Example 2-1: Adding Single Geographic Area to DataBasket

A single geographic area for the variable “NAICS code” within the 2007 County Business Patterns dataset is added by following the steps below.

Example 2-1: Selecting Counties for the County Business Patterns Dataset		
Steps	What to do	Why do it?
1	Select the 2007 instance of the County Business Patterns county level dataset. View the “2007 CBP County Variables,” select “NAICS code” and add it to the DataBasket. (For detailed instructions on these actions, see <i>Introduction, Example 1: Choose the Dataset and the Variables You Want to View.</i>)	This is a typical sequence for adding variables to the DataBasket. An alternative method would be to select the geographic areas first.

2	In the “Required Variables” window that opens after you have added “NAICS code” to the DataBasket, select the variable “Selectable Geographies” by double-clicking its row.	This action opens the “Ferrett Geography Codebook.”
3.1 3.2 3.3	In the left-hand panel of the “Ferrett Geography Codebook”, highlight “County” (3.1) and “FIPS State Code > FIPS County Code” under “Hierarchies” (3.2). Click “Use Hierarchy” (3.3).	This defines the geographic areas for the variables added to the DataBasket.
4.1 4.2 4.3	Select “New Jersey” from the list of states (4.1), hold down the mouse (or double-click) to drag it into the center panel (4.2), highlight it and then click “Next Level” (4.3, see Figure 2-16). Figure 2-17 shows the result of these	This action displays all counties in New Jersey.
5.1 5.2 5.3	In the left panel, highlight, drag and drop “Select All” to the right panel (5.1, 5.2). The complete list of counties will be displayed in a new window (Figure 2-19); click on “Finish” (5.3) to add the geography variable to the DataBasket.	This action places all geographies into your DataBasket.

Step 2



21 Variables returned from search. 2 variables selected in DataBasket.				
REQUIRED VARIABLES				
Selected	Topic	Name	Availability	Variable Label
<input checked="" type="checkbox"/> Needs to be selected	Selectable Geographies	Geography	1998 - current	Geographic Items
<input checked="" type="checkbox"/> Values already selected	CBP County Level Variables	naics	1998 - 2007	NAICS Code

Figure 2-14: Prompt to Select Geographic Variables

Step 3

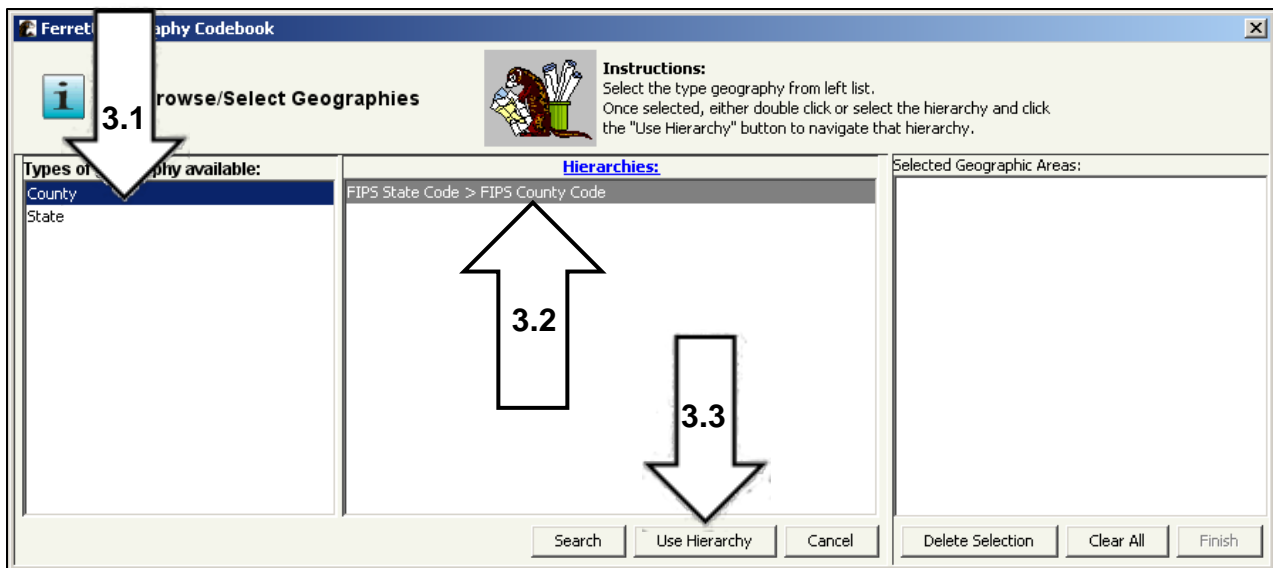


Figure 2-15: Select FIPS County Code

Step 4

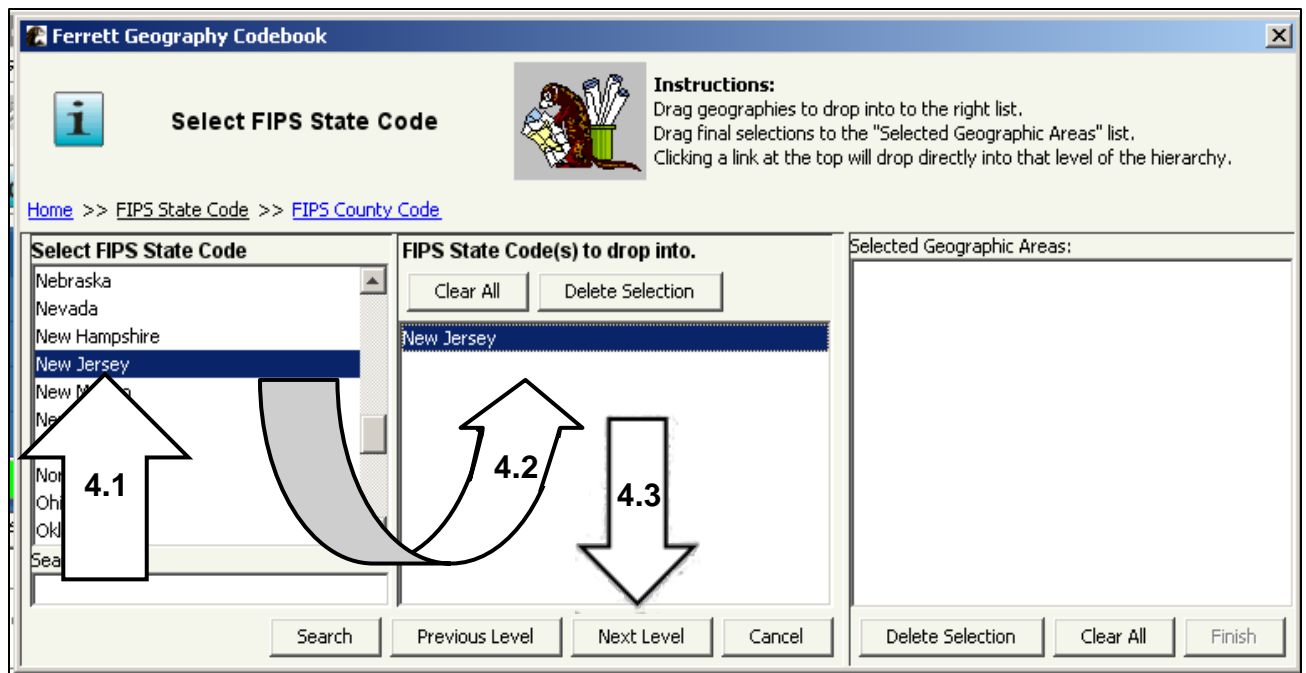


Figure 2-16: Select and Drag New Jersey from FIPS State Code List

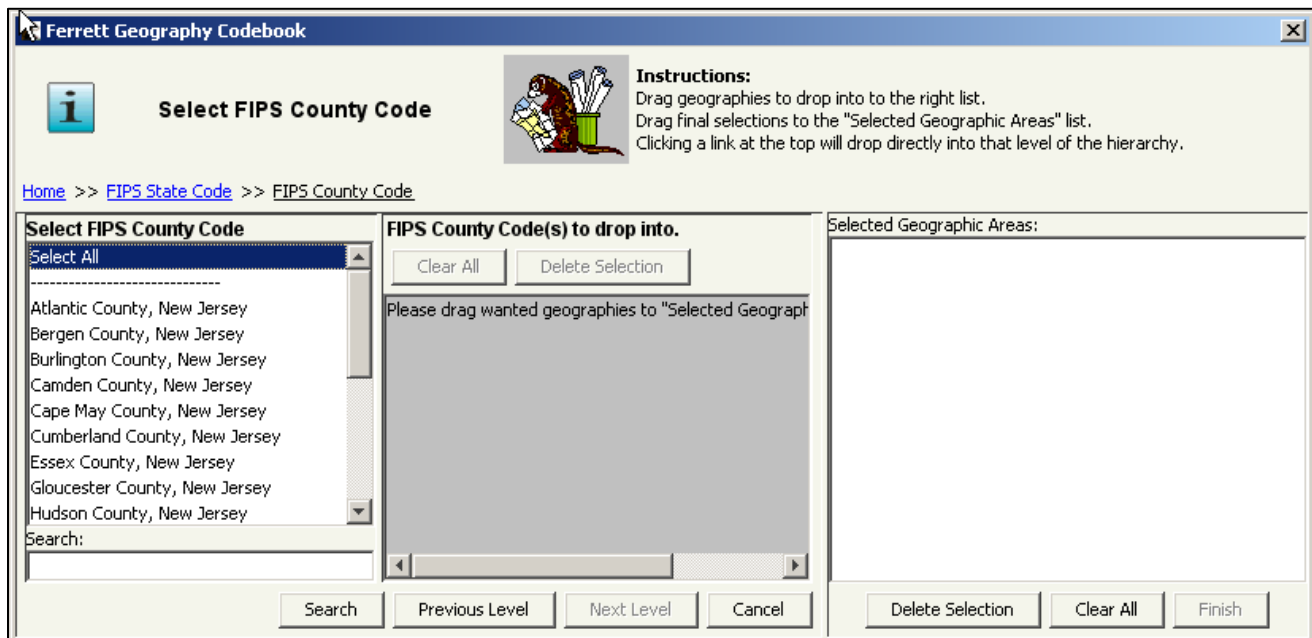


Figure 2-17: After Clicking “Next Level”

Step 5

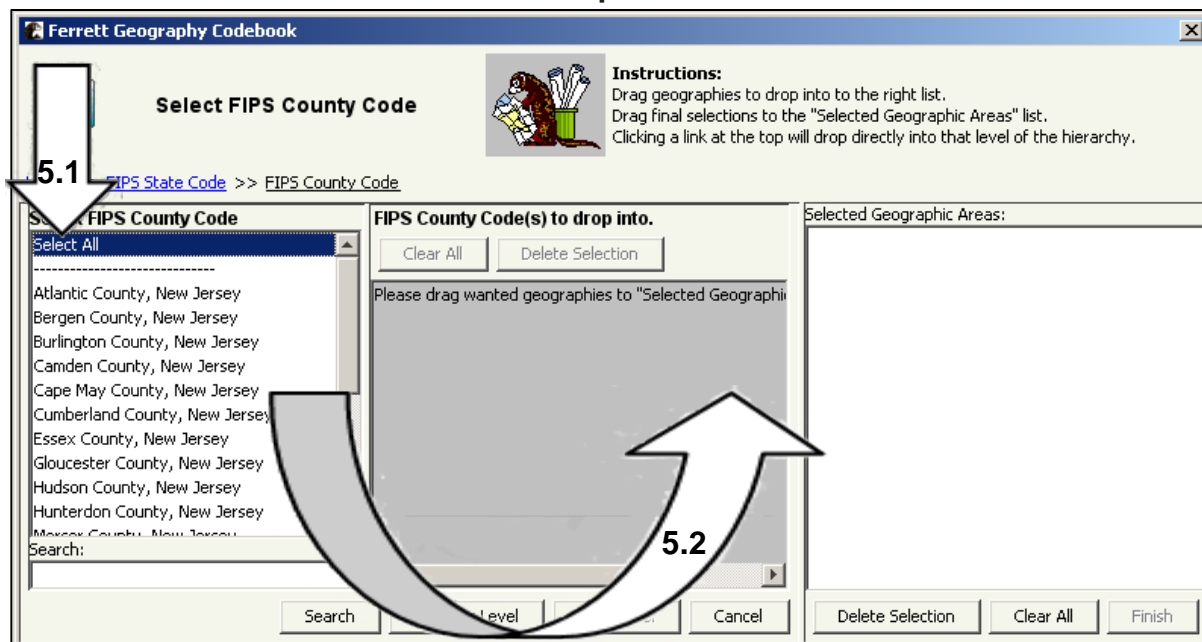


Figure 2-18: Drag “Select All” into Right Panel

Step 5 (con't.)

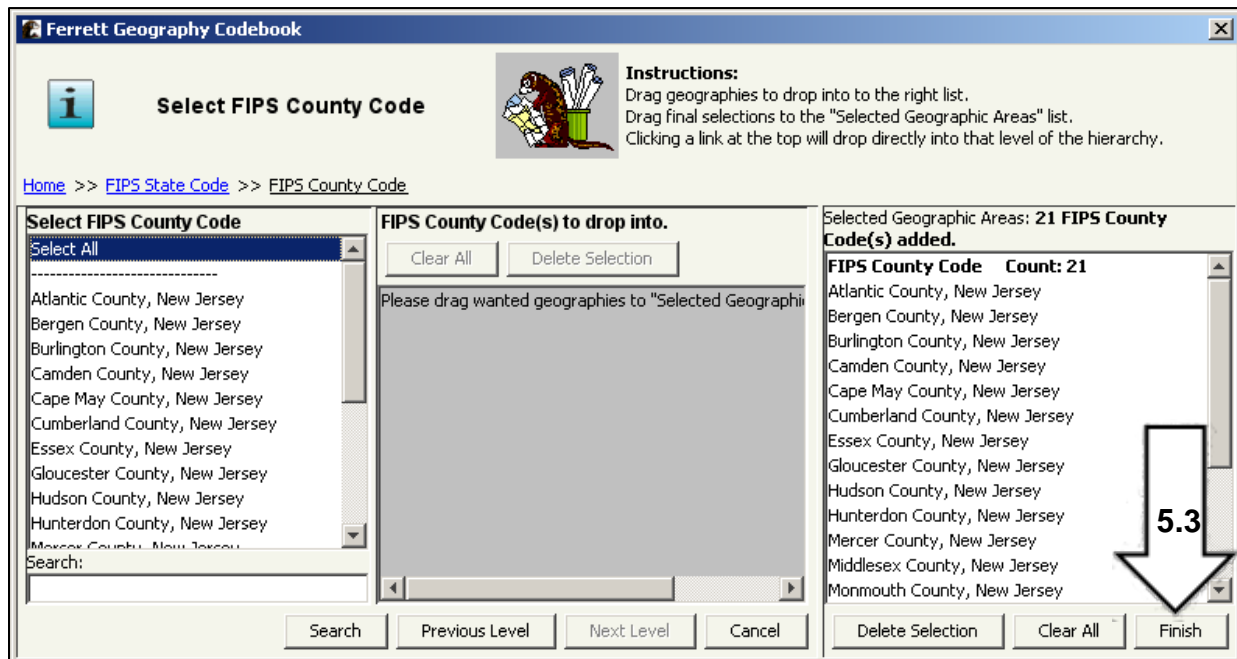


Figure 2-19: Click on “Finish”

Example 2-2: Adding Multiple Geographic Areas to the DataBasket

This example demonstrates the selection of areas within a chosen hierarchy and the selection of a particular geographic area by its name. The city of Glen Burnie, Maryland and all of the counties in the state of Maryland will be selected and added to the DataBasket. Sex by Educational Attainment for the Population 25 Years and Over within the 2006-2010 Summary File of the American Community Survey (ACS) is the specific variable searched by following the steps below.

Example 2-2: Selecting Various Geographic Areas from 2006-2010 ACS		
Steps	What to do	Why do it?
1	After selecting the variables, a warning message appears, followed by the “Required Variables” list. Double-click the row “Needs to be selected” marked with a red “X.”	This action opens up the “Ferrett Geography Codebook.”
2.1 2.2 2.3	Highlight “County” (2.1) under “Types of geography available.” The “State (FIPS Code)>County” will appear in the center panel (2.2). Highlight it and click on “Use Hierarchy” (2.3) to continue.	This will bring you to the list of state FIPS codes.
3.1 3.2 3.3	Select “Maryland” from the list of states, drag it into the second panel (3.2) and then click “Next Level” (3.3).	This action will bring you to the list of places within the state of Maryland.

4.1 4.2 4.3	After selecting “Maryland,” there will be a listing of all places in Maryland. Highlight “Select All” (4.1) and drag it to “Selected Geographic Areas” (4.2). Once all the counties are shown in “Selected Geographic Areas,” click on “Home” in the breadcrumb trail toward the upper left of the window.	You must place all the desired geographic areas into the “Selected Geographic Areas” panel in order to search within those areas. Clicking “Home” allows you to begin a new search while keeping all the geographic areas you selected.
5.1 5.2	Click on the “Search” button to bring up the search bar/window, type “Burnie” in the “Search” window (5.1) and then click “Search” (5.2).	This action searches for all areas within the Maryland places with the word “Burnie” in it.
6.1 6.2 6.3	Select “Glen Burnie CDP, Maryland” in the “Results” window (6.1) and drag it into the “Selected Geographic Areas” window. Then, scroll to the bottom of the list in “Selected Geographic Areas” to display “Glen Burnie CDP, Maryland” (6.2) then click “Finish” (6.3, see Figures 2-25 and 2-26).	You now have Glen Burnie, Maryland added to your DataBasket.

Step 1

Highlight the variables you are interested in

Currently viewing: B15002. Sex by Educational Attainment for the Population 25 Years and over 35 Va

3 variables selected in DataBasket. Browse/Select Highlighted

Selected	Topic	Name	Availability	Variable Label
<input checked="" type="checkbox"/> Default value selected for you	Selectable Geographies	COMPONENT	2005-2009 - current	Geographic Component
<input type="checkbox"/> Needs to be selected	Selectable Geographies	Geography	2005-2009 - current	Geographic Items

Figure 2-20: Required Variables Message

Step 2

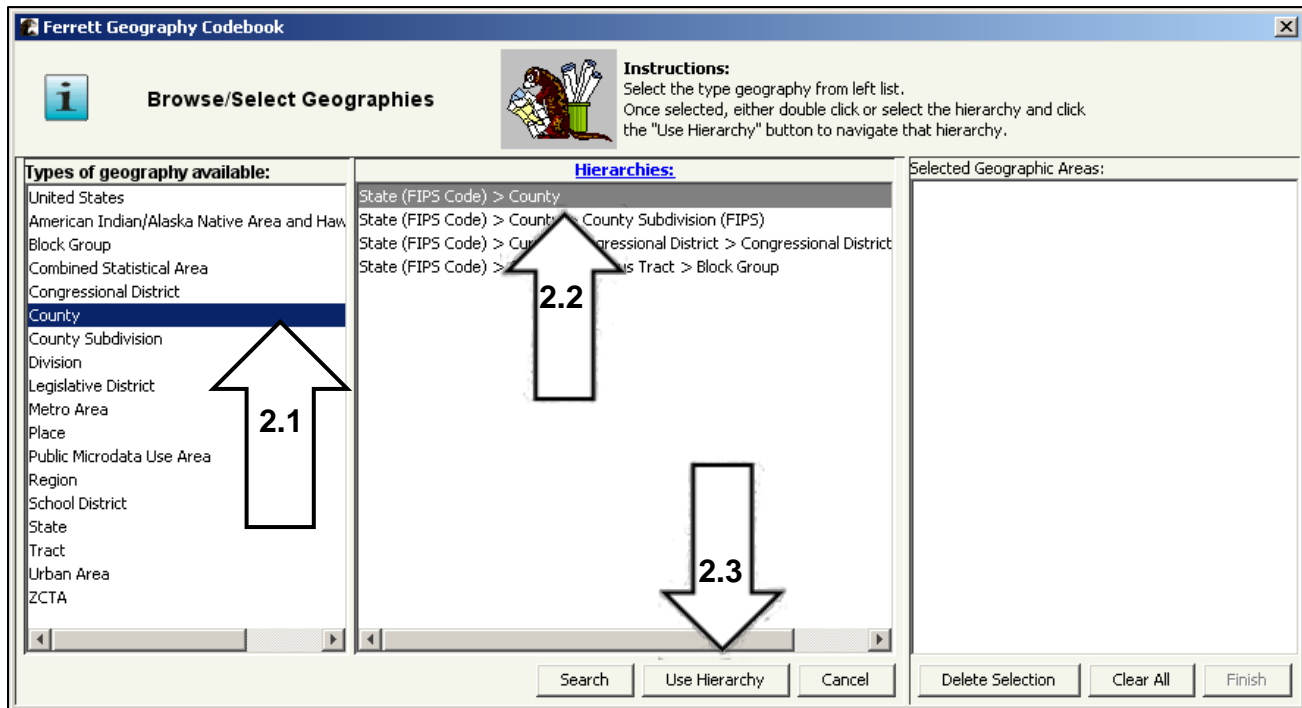


Figure 2-21: Geography Hierarchy Selected

Step 3

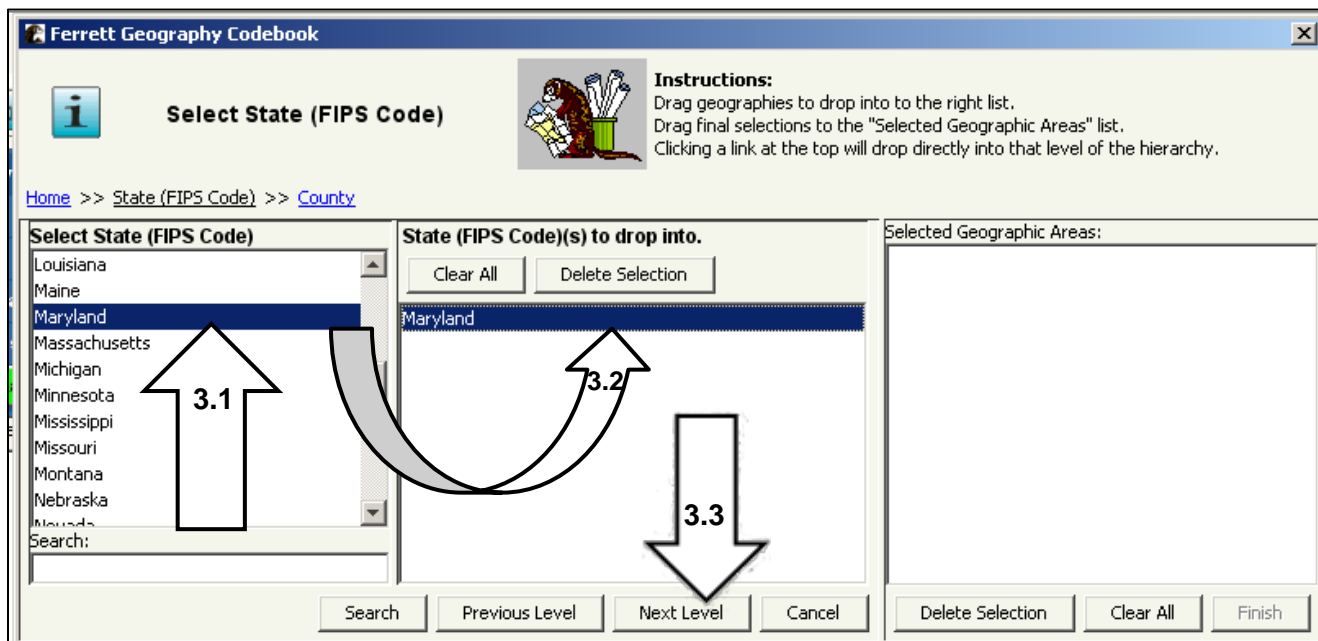


Figure 2-22: State of Maryland Selected

Step 4

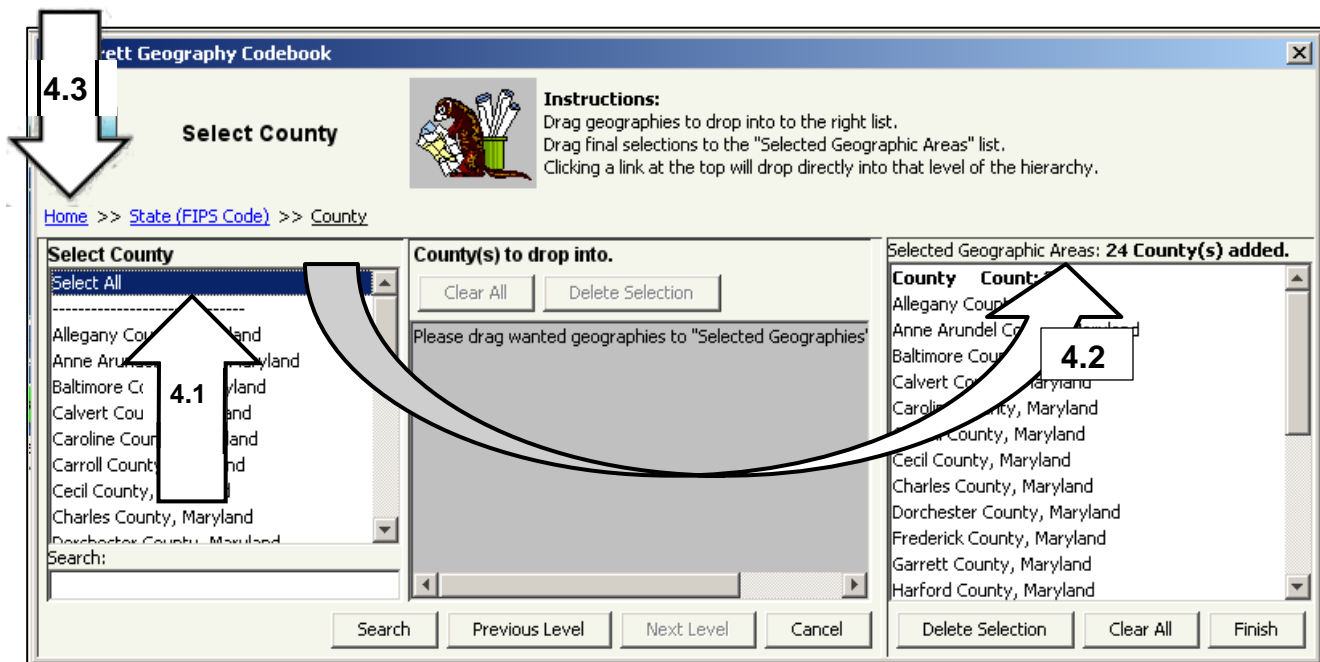


Figure 2-23: All Counties in Maryland Added to Selected Geographic Areas

Step 5

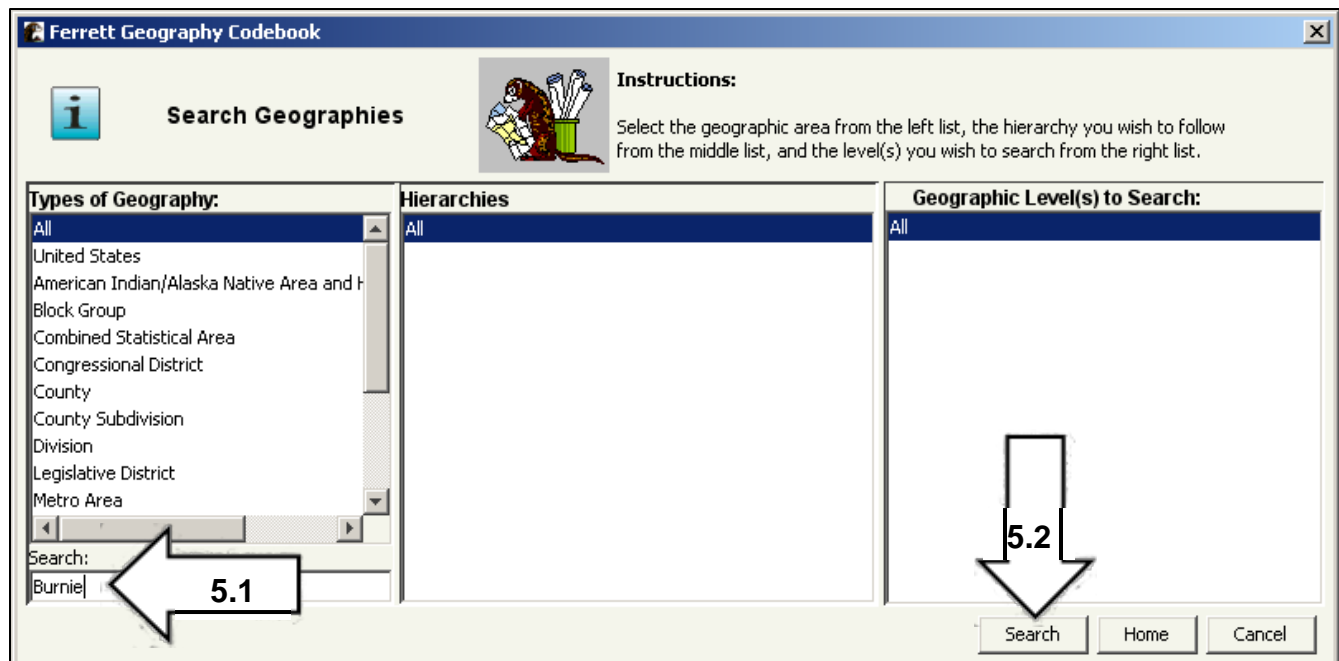


Figure 2-24: Search Parameter Defined as "Burnie"

Step 6

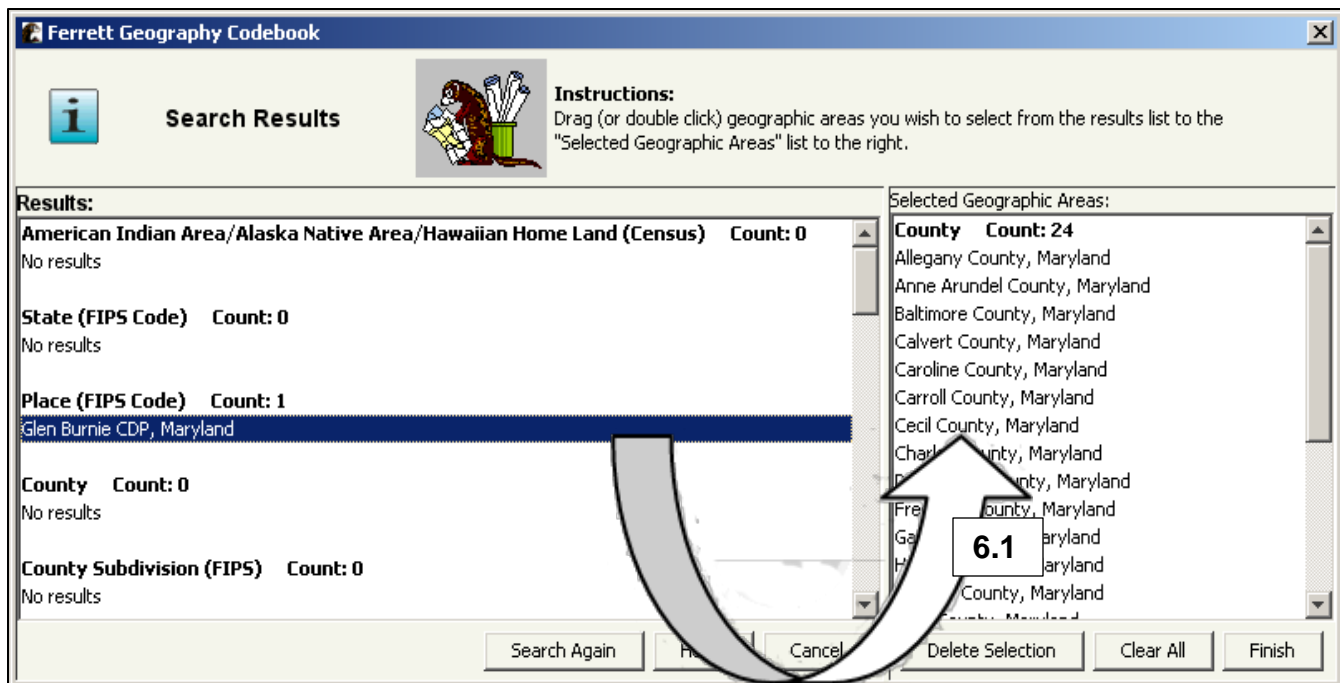


Figure 2-25: “Glen Burnie” Added to “Selected Geographic Areas”

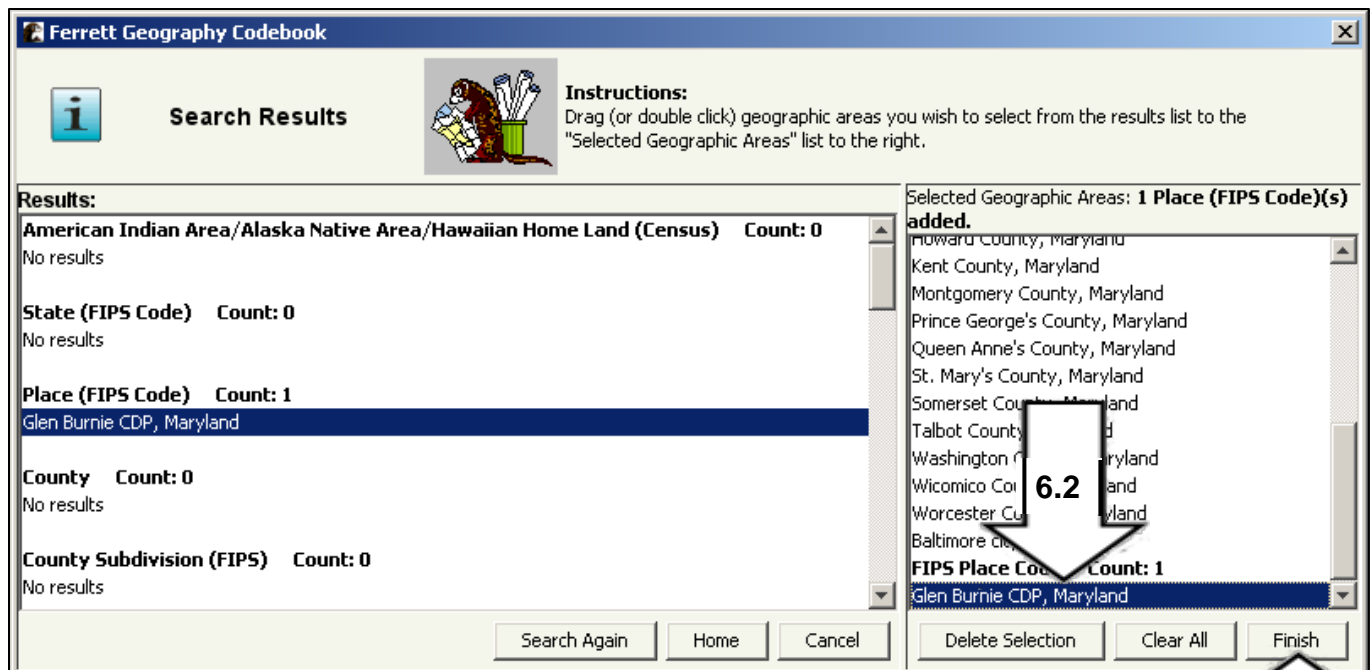


Figure 2-26: “Glen Burnie” Added to DataBasket

6.3

CHAPTER 3: RECODES & OTHER OPTIONS

What is a Recode?

A recode is a new variable that you create by combining or collapsing the value categories of an existing variable. For example, a recode of an age variable with values ranging from zero to 80 might have four age groups of 20 years each (0-20, 21-40, 41-60, 61-80). (DataFerrett allows you to define the groups any way that you like.) The new variable with four values is called a “Recode variable” which is added to the DataBasket once it is defined.

The Recode feature allows you to group values of variables together in several different ways. This chapter uses three examples to demonstrate the capabilities of the recode feature.

NOTE: Currently you can only create a recode variable when using a microdata dataset.

Example 3-1: Creating a Recode Variable

We begin by continuing the example from the Introduction, resuming at Step 4 of Task 2. The steps below will demonstrate how to create a simple recode from a variable with a large range of values.

Example 3-1: Creating a Recode Variable		
Steps	What to do	Why do it?
1.1 1.2	Highlight the variable "Education Assistance Amount - Family" (1.1) in the DataBasket list on the "Step 2" screen and then click "Recode Variable" on the right (1.2).	A recode variable can only be based on a variable in the DataBasket.
2.1 2.2 2.3	Type "Ed assistance amt" in the blank space (2.1) in the upper-left corner of the Recode window. Go to the "continuous values" section, type in 10000 (2.2), then click "Recode" (2.3).	This action sets the initial range value (value 1) at 1-10,000 in order to facilitate presentation and analysis of the data.

3	The recode window now shows two values: value 1 with a range of 1-10,000 and value 2 with a range of 10,001-3,899,961. Enter a new range end point of 30,000 for value 2 in the “continuous values” section (3.1) and click “Recode” (3.2).	These actions set the range for value 2 at an interval of 20,000 to further facilitate presentation and analysis of the data.
4	Continue adding additional ranges at intervals of 20,000 each until you reach an end point of 90,000 (clicking “Recode” after each interval.)	These actions set the range for values 3-5 at intervals of 20,000 to further facilitate presentation and analysis of the data.
5	Enter a new range end point of 120000 for value 6 in the “continuous values” section and click “Recode.” Click “Recode” again to establish value 7 with a range of 120001 to 3899961, then click “Ok”.	These actions define the final range values.
6	Figure 3-7 shows the new recode variable.	The recode variable, RECODE1, has now been added to the DataBasket.

Step 1

Introduction Step1: Select Dataset & Variable Step2: DataBasket/Download/Make A Table

Review your variables then go back to select more variables or go on to get data

Download Make A Table

1.2

Act on Your Query: —

Recode Variable(s)

Delete Variable(s)

View/Modify Variable(s)

Advanced Sql Option

Change Longitudinal Period

Add TimeSeries Time

Merge Datasets

Save Selected Variable(s) CodeBook

Create Multi-Variable Data Step

Current Query Variables: —

Name	Variable Label	Availability
FEDVAL	Educational assistance amount - Family Mar 1992 - current	

1.1

Figure 3-1: Select Variable and Begin Recode

Step 2

2.1 assistance amt is label for the Variable Recode of FEDVAL

Highlight the value(s) to recode/regrup

Value	Description
0	None or not in universe

Set to value 1

continuous values

Within the range from 1 through

Either:

Set to value 1

Or:

Subgroups repeat by

(e.g. by 10=0-9,10-19,...)

Label	Values
1 Not Elsewhere Classified (nec.)	{0 ,Between 1 and 3899961 }

Redefine Current Row

Figure 3-2: Recode Definition

Step 3

assistance amt is label for the Variable Recode of FEDVAL

Highlight the value(s) to recode/ regroup

Value	Description
0	None or not in universe

Set to value 2

continuous values

Within the range from 10001 through

Either:

Set to value 2

Or:

Subgroups repeat by

(e.g. by 10=0-9,10-19,...)




Label	Values
1 Between 1 and 10000	{Between 1 and 10000}
2 Not Elsewhere Classified (nec.)	{0 ,Between 10001 and 3899961 }

Redefine Current Row

Figure 3-3: Recode Range Interval Modification

Step 4

Ferrett Microdata Recode1

 **Recode/Regroup Variables**  

is label for the Variable Recode of FEDVAL

Highlight the value(s) to recode/regroup

Value	Description
0	None or not in universe

Set to value

continuous values

Within the range from 90001 through

Either:

Set to value

Or:

Subgroups repeat by

(e.g. by 10=0-9,10-19,...)

Label	Values
1 Between 1 and 10000	{Between 1 and 10000}
2 Between 10001 and 30000	{Between 10001 and 30000}
3 Between 30001 and 50000	{Between 30001 and 50000}
4 Between 50001 and 70000	{Between 50001 and 70000}
5 Between 70001 and 90000	{Between 70001 and 90000}
6 Not Elsewhere Classified (nec.)	{0 ,Between 90001 and 3899961 }

Figure 3-4: Continuation of Recode Range Interval Modification

Step 5

Ferrett Microdata Recode1

Recode/Regroup Variables

assistance amt is label for the Variable Recode of FEDVAL

Highlight the value(s) to recode/regroup

Value	Description
0	None or not in universe

Set to value **6** **Recode**

continuous values

Within the range from 90001 through **120000**

Either:

Set to value **6** **Recode**

Or:

Subgroups repeat by

(e.g. by 10=0-9,10-19,...) **Recode**

Label	Values
1 Between 1 and 10000	{Between 1 and 10000}
2 Between 10001 and 30000	{Between 10001 and 30000}
3 Between 30001 and 50000	{Between 30001 and 50000}
4 Between 50001 and 70000	{Between 50001 and 70000}
5 Between 70001 and 90000	{Between 70001 and 90000}
6 Not Elsewhere Classified (nec.)	{0 ,Between 90001 and 3899961 }

Redefine Current Row

Ok **Cancel**

Figure 3-5: Continuation of Recode Range Interval Modification

Step 5, cont.

Recode/Regroup Variables

Tell me about

Things on the screen

Ed assistance amt

 is label for the Variable Recode of FEDVAL

Highlight the value(s) to recode/regroup

Value	Description
0	None or not in universe

Set to value 7

Recode

continuous values

Within the range from 120001 through 3899961

Either:

Set to value 7

Recode

Or:

Subgroups repeat by

(e.g. by 10=0-9,10-19,...)

Recode

Label	Values
1 Between 1 and 10000	{Between 1 and 10000}
2 Between 10001 and 30000	{Between 10001 and 30000}
3 Between 30001 and 50000	{Between 30001 and 50000}
4 Between 50001 and 70000	{Between 50001 and 70000}
5 Between 70001 and 90000	{Between 70001 and 90000}
6 Between 90001 and 120000	{Between 90001 and 120000}
7 Not Elsewhere Classified (nec.)	{0 ,Between 120001 and 3899961 }

Redefine Current Row

Ok

Cancel

Figure 3-6: Recode Variable Completely Defined

Step 6

Review your variables then go back to select more variables or go on to get data

Download

Make A Table

Current Query Variables from CPS (March Supplement):

Name	Variable Label	Availability
FEDVAL	Educational assistance amount - Family	Mar 1992 - current
RECODE1	Ed assistance amt	Mar 1992 - current

Figure 3-7: New Recode Variable

Other Features of the Recode Window

A continuous range variable is one in which possible values are defined only as a range; in other words, each value is not labeled or defined. The previous example dealt with such variables. Age and income are often continuous range variables because they cover a broad range of values. In the following example, we will demonstrate how to use the "Subgroups repeat by" feature of the recode window to create equal-sized value groupings. Figure 3-8 shows the recode window for an age (continuous range) variable, using an age variable from the Public Use Microdata Sample (PUMS) of the American Community Survey.

Example 3-2: Using the "Subgroups repeat by" Feature

To begin, place the variable "PUMS Age" into your DataBasket.

.Example 3-2: Using the "Subgroups repeat by" Feature		
Steps	What to do	Why do it?
1.1 1.2	Highlight the "PUMS Age" variable (1.1) and then click the "Recode Variable" button (1.2).	A recode variable is based on a variable in the DataBasket.
2.1 2.2	For the first recode value, you will use an age range between 0-17 (2.1), which will leave you with a remaining continuous value between 18 and 99 (2.2).	Dividing a large range of values into smaller ranges makes the values easier to work with.
3.1 3.2	Enter values in increments of 20 by typing '20' in the "Subgroups repeat by" category (3.1) then click "Recode" (3.2).	Using the "Subgroups repeat by" category allows you to choose an interval number that will automatically generate a complete listing of recode values.
4	Because the last age range variable consists of only two years, the recode variables are not evenly distributed.	The maximum number you assign the last value determines how the age ranges will be divided.
5	Figure 3-12 shows how you can more evenly distribute the values by setting the maximum age range to 77 instead of 99.	Recode variables are now more evenly distributed.
6	Figure 3-13 shows the final recode with the age range of 78 to 99.	This changes the label of the final variable from "Not elsewhere classified" to the range of values.

Step 1

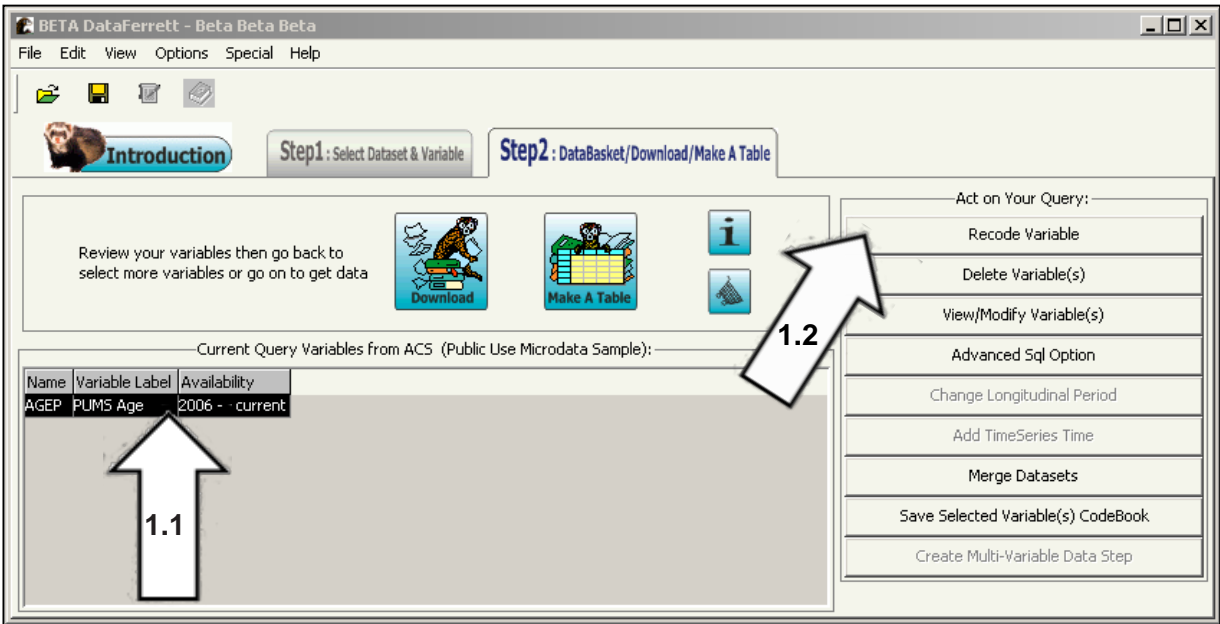


Figure 3-8: Begin Recode of PUMS Age Variable

Step 2

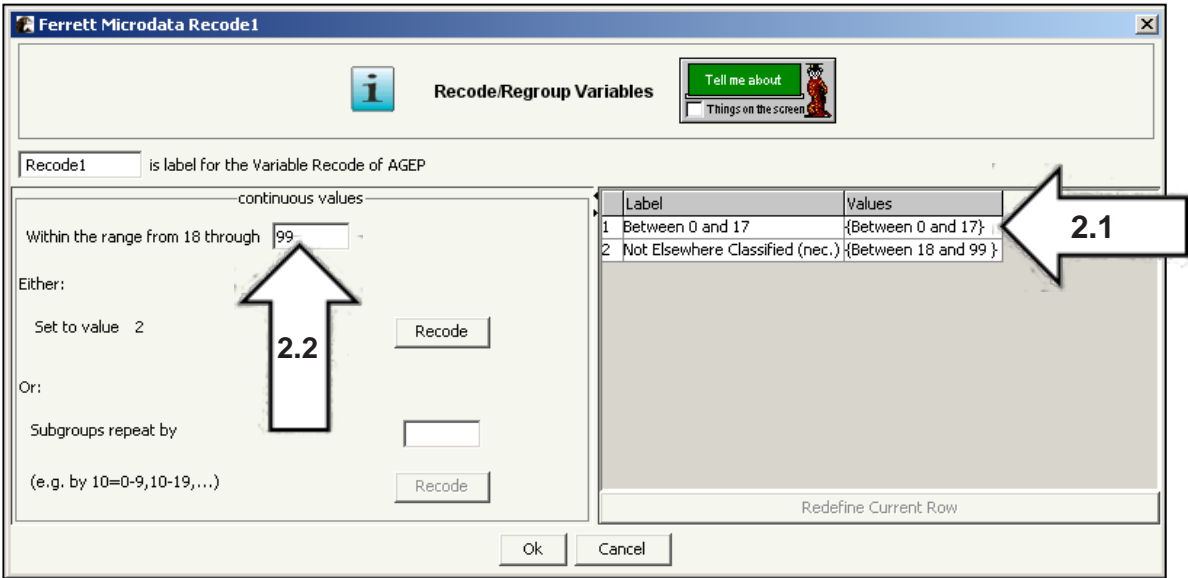


Figure 3-9: Results for Recode Variable by Age

Step 3

The screenshot shows the 'Ferrett Microdata Recode3' dialog box. The title bar is 'Ferrett Microdata Recode3'. The main area has a header 'Recode/Regroup Variables' with an information icon and a 'Tell me about' button. Below the header, it says 'Recode3 is label for the Variable Recode of AGE'. The 'continuous values' section has a range from 18 to 99. The 'Either:' section has 'Set to value 2' and a 'Recode' button. The 'Or:' section has 'Subgroups repeat by' with a value of 20 and a 'Recode' button. A large arrow labeled '3.1' points to the '20' value. Another arrow labeled '3.2' points to the 'Recode' button in the 'Or:' section. On the right, there is a table with two columns: 'Label' and 'Values'.

Label	Values
1 Between 0 and 17	{Between 0 and 17}
2 Not Elsewhere Classified (nec.)	{Between 18 and 99}

At the bottom right, there is a 'Redefine Current Row' button. At the bottom left, there are 'Ok' and 'Cancel' buttons.

Figure 3-10: Using the "Subgroups repeat by" Feature

Step 4

The screenshot shows the 'Ferrett Microdata Recode2' dialog box. The title bar is 'Ferrett Microdata Recode2'. The main area has a header 'Recode/Regroup Variables' with an information icon and a 'Tell me about' button. Below the header, it says 'Recode2 is label for the Variable Recode of AGE'. The 'continuous values' section has a range from 0 to 99. The 'Either:' section has 'Set to value 6' and a 'Recode' button. The 'Or:' section has 'Subgroups repeat by' with a value of 6 and a 'Recode' button. A large arrow labeled '4' points to the 'Recode' button in the 'Or:' section. On the right, there is a table with two columns: 'Label' and 'Values'.

Label	Values
1 Between 0 and 17	{Between 0 and 17}
2 Between 18 and 37	{Between 18 and 37}
3 Between 38 and 57	{Between 38 and 57}
4 Between 58 and 77	{Between 58 and 77}
5 Between 78 and 97	{Between 78 and 97}
6 Between 98 and 99	{Between 98 and 99}

At the bottom right, there is a 'Redefine Current Row' button. At the bottom left, there are 'Ok' and 'Cancel' buttons.

Figure 3-11: Recode Variables by Intervals

Step 5

Recode2 is label for the Variable Recode of AGEF

continuous values

Within the range from 18 through 77

Either:

Set to value 2

Recode

Or:

Subgroups repeat by 20

(e.g. by 10=0-9,10-19,...)

Recode

Label	Values
1 Between 0 and 17	{Between 0 and 17}
2 Not Elsewhere Classified (nec.)	{Between 18 and 99 }

Redefine Current Row

Ok Cancel

Figure 3-12: Alternative Approach to Setting Up Recode Value

Step 6

Recode3 is label for the Variable Recode of AGEF

continuous values

Within the range from 78 through 99

Either:

Set to value 5

Recode

Or:

Subgroups repeat by

(e.g. by 10=0-9,10-19,...)

Recode

Label	Values
1 Between 0 and 17	{Between 0 and 17}
2 Between 18 and 37	{Between 18 and 37}
3 Between 38 and 57	{Between 38 and 57}
4 Between 58 and 77	{Between 58 and 77}
5 Not Elsewhere Classified (nec.)	{Between 78 and 99 }

Redefine Current Row

Ok Cancel

Figure 3-13: Results before Selecting “Recode” for New Final Age Range, 78 to 99

Editing Labels for Recode Variables

Notice how the “Not Elsewhere Classified (nec)” label is still used for the last value for the recode variable in Figure 3-13. You can change the label for this recode value by double clicking on the label and changing it to something consistent with the other labels, e.g., “Between 78 and 99” or

“78 and older.” You can also change the label of the recode variable itself by clicking in the box containing the default name (“Recode3” in Figure 3-13) and changing the label to something more meaningful.

NOTE: Once you click on the “OK” button for a recode variable, the labels for the variable or any of its values can only be changed by using the “View/Modify” button under “Act on your Query”.

The example above deals with the "continuous range" variables, i.e., variables that have large ranges of values. Each value within a range is exactly what the variable represents; for example, a value of three for age means the person is a three-year old child. However, many variables in a dataset may have relatively few values, and each value has a specific meaning that can only be understood by examining the definition (“label”) of the value. These variables are called “categorical variables.” Creating a recode variable based on a categorical variable enables regrouping the original variable’s values into fewer categories.

For example, the ACS PUMS variable MAR (marital status) variable has the values:

- Married
- Widowed
- Divorced
- Separated
- Never Married

Recoding categorical variables such as this one is demonstrated in the example below.

Example 3-3: Recoding a Categorical Variable

To begin this example, go to the American Community Survey and select any year in the "Public Use Microdata Sample," then click on "Population" in variables to search. You will be recoding the "Marital Status" variable.

Example 3-3: Recoding a Categorical Variable		
Steps	What to do	Why do it?
1.1 1.2	Select the "Marital Status" (1.1) from the list of variables in the DataBasket. With the variable label highlighted, click on the "Recode Variable" (1.2) button.	You will be creating recode variables to reduce the number of variables to "Married" and "Not Married."
2.1 2.2	The initial recode window for this variable is shown in Figure 3-15. To create a new recode variable for marital status, highlight “Married” (2.1) and click	The first recoded variable, "Married," is created.

3.1 3.2	Combine the rest of the values into one value labeled "Not Married" by highlighting all four values (3.1) and clicking "Recode" (3.2).	This allows you to create the new variable. You have now created two recodes from five values by limiting categories. Note that the new recode is labeled "RecodeValue_2"
4	To rename the new recode, fill in the label to the left of the "Recode/Regroup Variables" window. To do this, double click in the label field and highlight the existing label, then type in the new label name "Not married (incl widowed and	The two categories, "Married" and "Not married", are now properly labeled.

Step 1

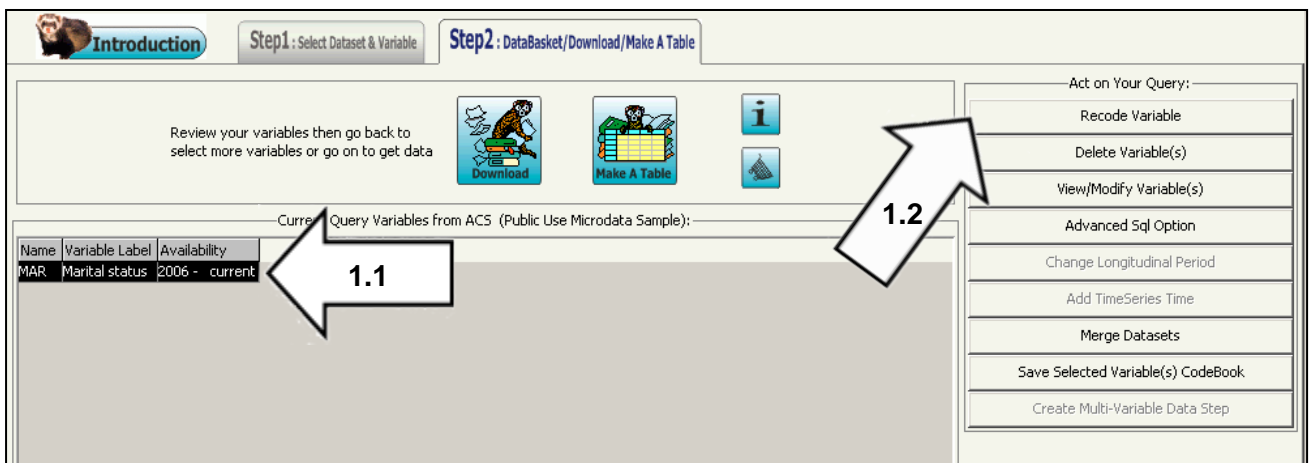


Figure 3-14: Recoding "Marital Status" Variable

Step 2

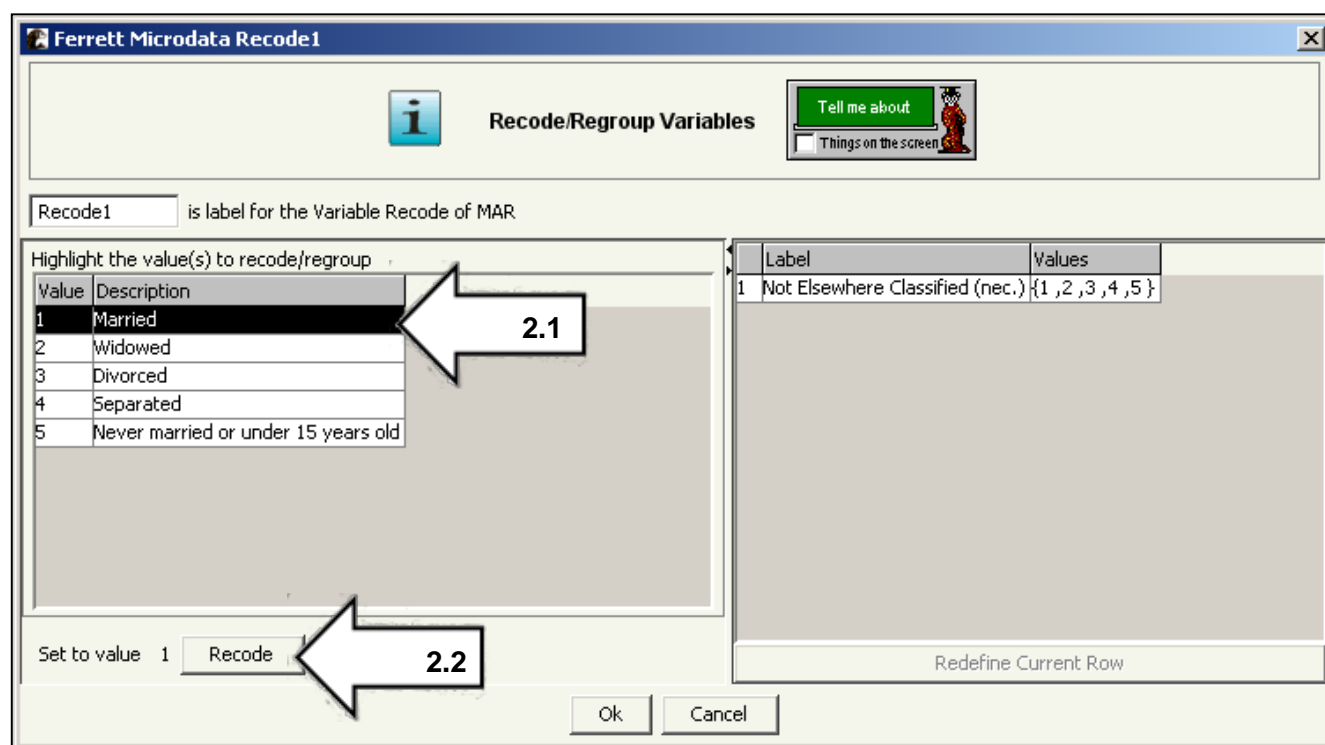


Figure 3-15: Initial Recode Window for Variable

Step 3

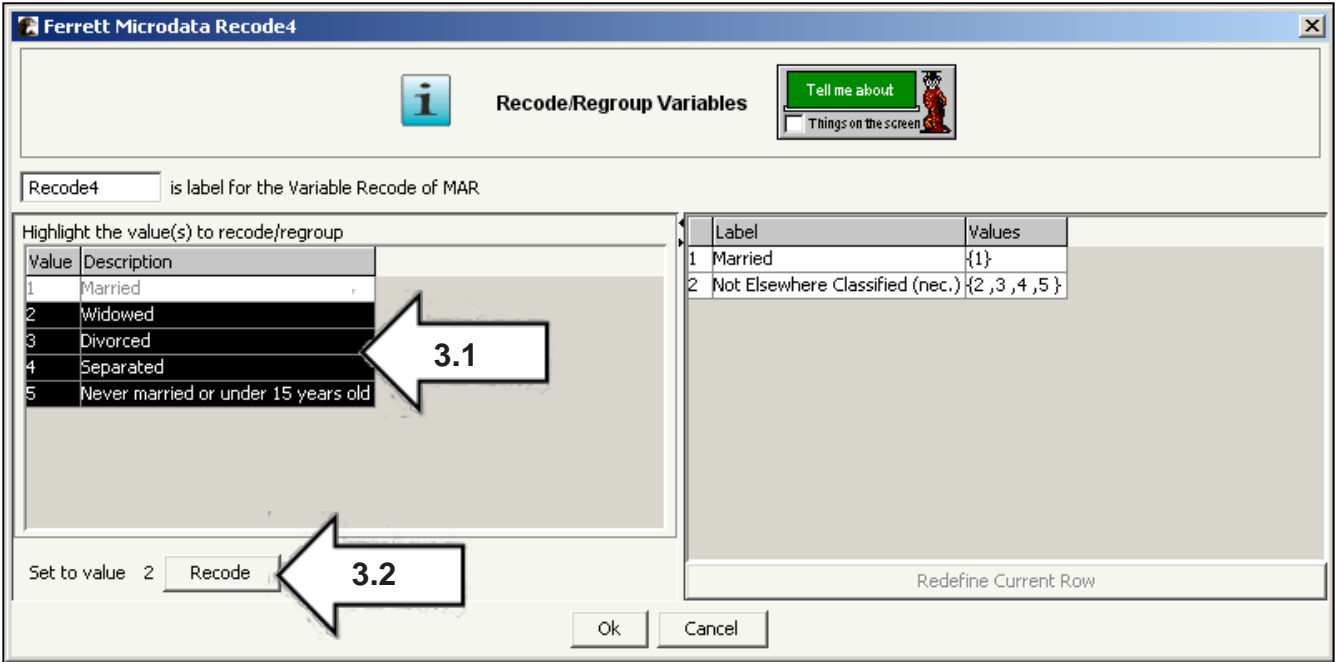


Figure 3-16: Select Items to be Grouped

Step 3 (cont.)

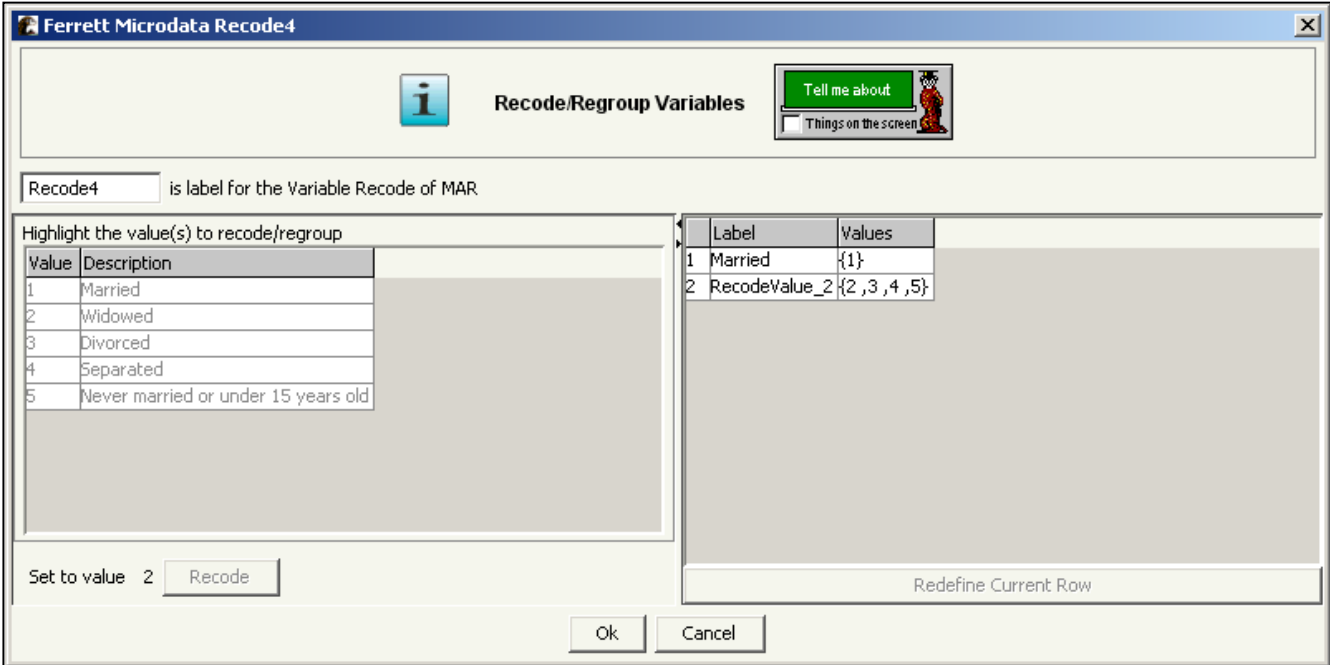


Figure 3-17: Results after Grouping Variables

Step 4

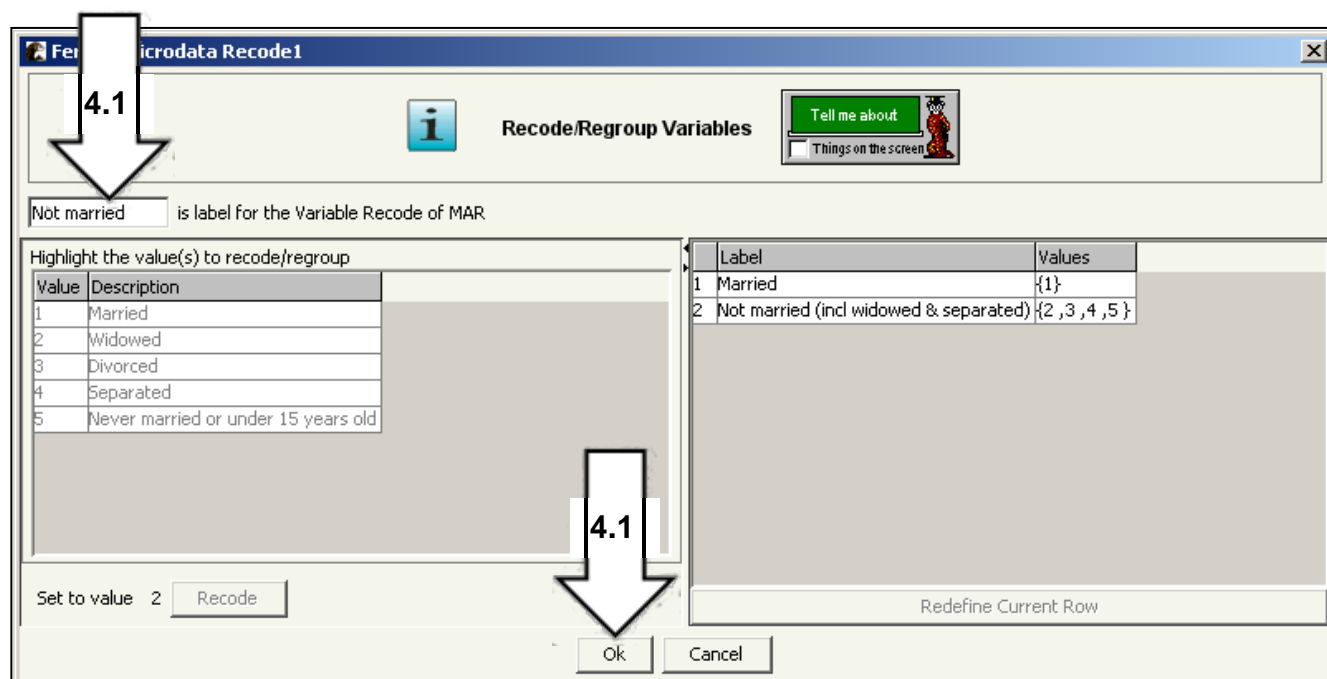


Figure 3-18: Renaming New Recode Variable

Other Options in Step 2

The “Act on Your Query” list in “Step 2” offers many choices in addition to “Recode Variable(s)”. Brief explanations of the basic options available are provided below.

Delete Variable(s)

You may decide that certain variables are not necessary to include with your current data. You can remove them from the DataBasket, especially if you plan to save all the variables in the DataBasket to a DataFerrett Session File (FSF) for later use. To use this feature, highlight the variable(s) you want to delete and click on the “Delete Variable(s)” button on the “Act on Your Query” list. Exercising this function removes variables from the DataBasket, an action that cannot be reversed. You must exercise care in choosing the variable(s) to delete to avoid unintentionally removing a variable that is required.

View/Modify Variable(s)

This feature allows you to view all values for a variable and remove selected values. This is done in a manner similar to the way you browse variables and delete values in the Browse/Select Variables & Values window in "Step 1."

To use this feature, highlight the variable(s) in the DataBasket and click on the “View/Modify Variable(s)” choice on the “Act on Your Query” list. This opens the Browse/Select Variables & Values window where values for the highlighted variable can be deleted by selecting “Deselect All

Values” and clicking on the “OK” button. A message will appear confirming the number of variables that have been modified. You may confirm the modification by selecting "OK" or "Cancel" to proceed without the modifications.









Save Selected Variable(s) Codebook









This choice allows you to save the documentation for the variables in your DataBasket to an ASCII text or HTML file. The documentation includes the variable name, label and each value along with its description. This choice is always active in the "Step 2" screen, since it operates on the entire DataBasket. It is not possible to view the contents of the codebook from within DataFerrett.

CHAPTER 4: MAKE A TABLE

Once you have added the selected variables to your DataBasket, the “Step 2: DataBasket/Download/Make a Table” window provides the methodology required for organizing the data into a table. This methodology offers many options, detailed in “Ferrett Tabulation Tools” below, that allow you to organize and present your data in a number of ways such as statistical maps or graphs.

Ferrett Tabulation Tools

Toolbar Buttons: Descriptions & Functions		
Name	Icon	Description
The ‘i’ Button		Opens the table instruction box
Go Get Data		Returns the values for selected variables in the spreadsheet
Graph		Provides menu of graph options
Map		Provides option for creating a map from spreadsheet data; active only if a geographic variable that can be mapped has been selected and is defining the rows of the table
Point Map		Allows layer to display address points assuming that a dataset containing appropriate latitude and longitude information has been tabulated and is being added to an existing thematic map
The Time Series		Displays the data from selected cells in a table as a time series line graph for up to twelve time periods
Clear Spreadsheet		Removes all variables and values from an existing spreadsheet
Sort Ascending		Sorts a selected column or row in ascending order

Sort Descending		Sorts a selected column or row in descending order
Undo Sort		Returns table columns and rows to the original layout and data display
Spreadsheet Only		Displays all the columns in a table beyond what is seen by default in the table view; hides the list of selected variables displayed to the right of the table, the universe information displayed below the selected list of variables and the DataFerrett image, effectively showing as many columns as possible in the spreadsheet; returns to original table view when re-clicked
View Underlying		Displays the underlying data records that are available for selected variables NOTE: Aggregate (summed) data, recodes or other calculated data do not have underlying data records.
Decimal Format		Converts values in a selected column(s) or row(s) from an integer number to decimal values of up to three decimal places OR as numbers displayed in tens, hundreds, thousands, OR as a code formatted number; i.e. no commas
Show % of First Data Column		Converts all values in a table to percentages of the first column in a table; i.e., the denominator is the value in the cells of the first column in the table with the remaining cell values as the numerator
Show % of First Data Row		Converts all values in a table to percentages of the first row of the table; i.e., the denominator is the value in the cells of the first row in the table with the remaining cell values as the numerator
Show % of First Data Cell		Converts all values in a table to percentages of the value in the first data cell of the table; i.e., converting the values of the remaining cells in the table to percentages of the values in the first cell.


Tips for Modifying Table		Provides video tutorials demonstrating functions pertaining to modifying a table, including: <ul style="list-style-type: none"> • Adding variable rows • Replacing rows with a new variable • Nesting a variable • Adding pivots • Changing the geographic component of a table • Changing or modifying values • Recoding a variable • Modifying a variable • Deleting a value(s) from a selected variable (modifying the variable universe) • Creating a geographic pivot selection
--------------------------	---	--

Table 4-1: “Ferrett Tabulation Tools”

Make a Table: Basic Steps

There are basic steps and formats utilized in making a table within DataFerrett, described in the following sections. The options listed above are available only after these basic steps are completed.

Step 2: DataBasket/Download/Make a Table

The “Step 2” window tab is selected once all of the desired variables have been added to the DataBasket. Once selected, clicking on the “Make a Table” icon will open the “Ferrett Tabulation” screen in a new window. The Step 2 window, with the “Make a Table” icon highlighted and the Ferrett Tabulation screen are shown below.

Chapter 4: Make a Table



Figure 4-1: The “Step 2” Screen with “Make a Table” Highlighted

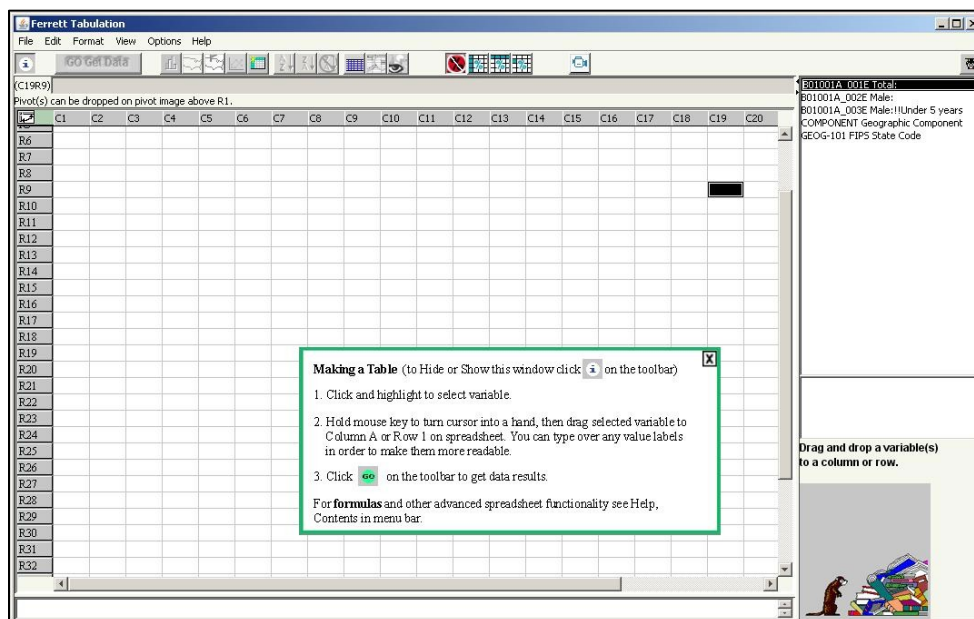


Figure 4-2: Ferrett Tabulation Screen

When opened initially, the Ferrett Tabulation Screen includes an abbreviated information window titled “Making a Table.” This window, shown below, may be hidden by clicking the Information icon on the toolbar or closed by clicking the “X” box in the upper right hand corner.

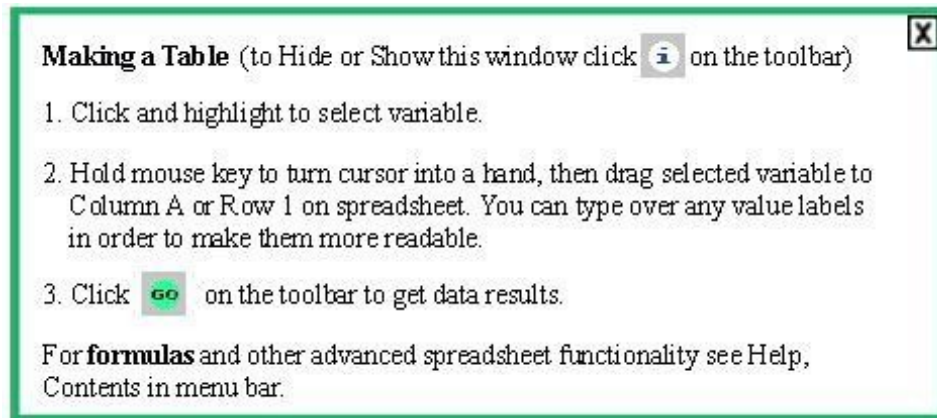


Figure 4-3: Making a Table Information Window

NOTE: The Ferrett Tabulation Screen may be minimized if a return to the Step 1: Select Dataset & Variables is needed to add additional variables to the DataBasket or to manipulate existing variables. The screen will be automatically updated to reflect changes to the selected variables.

The Ferrett Tabulation Screen

The Ferrett Tabulation toolbar, located at the top of the Ferrett Tabulation screen and shown below, contains tools to assist in the manipulation of data in the spreadsheet and the presentation of selected data. The toolbar also includes brief tutorials that describe how to perform various functions pertaining to modifying a table. A list of these tools, their icons and a brief description of each can be found in Table 4-1 at the beginning of this chapter.



Figure 4-4: Ferrett Tabulation Toolbar

The Ferrett Tabulation screen lists the variables in the DataBasket to the right of the tabulation grid. To begin populating the spreadsheet, highlight a variable (arrow 1, Figure 4-5), hold the mouse key down and drag it over to the first row/first column in the spreadsheet (cell R1, C1) (arrow 2, Figure 4-5). Repeat the process for the second variable, dragging it to cell R1, C2. You can multi-select variables by holding the shift key down, highlighting several variables at once and dragging them to the selected cell. Blank rows and/or columns may also be inserted between groups of variables if desired.

Chapter 4: Make a Table

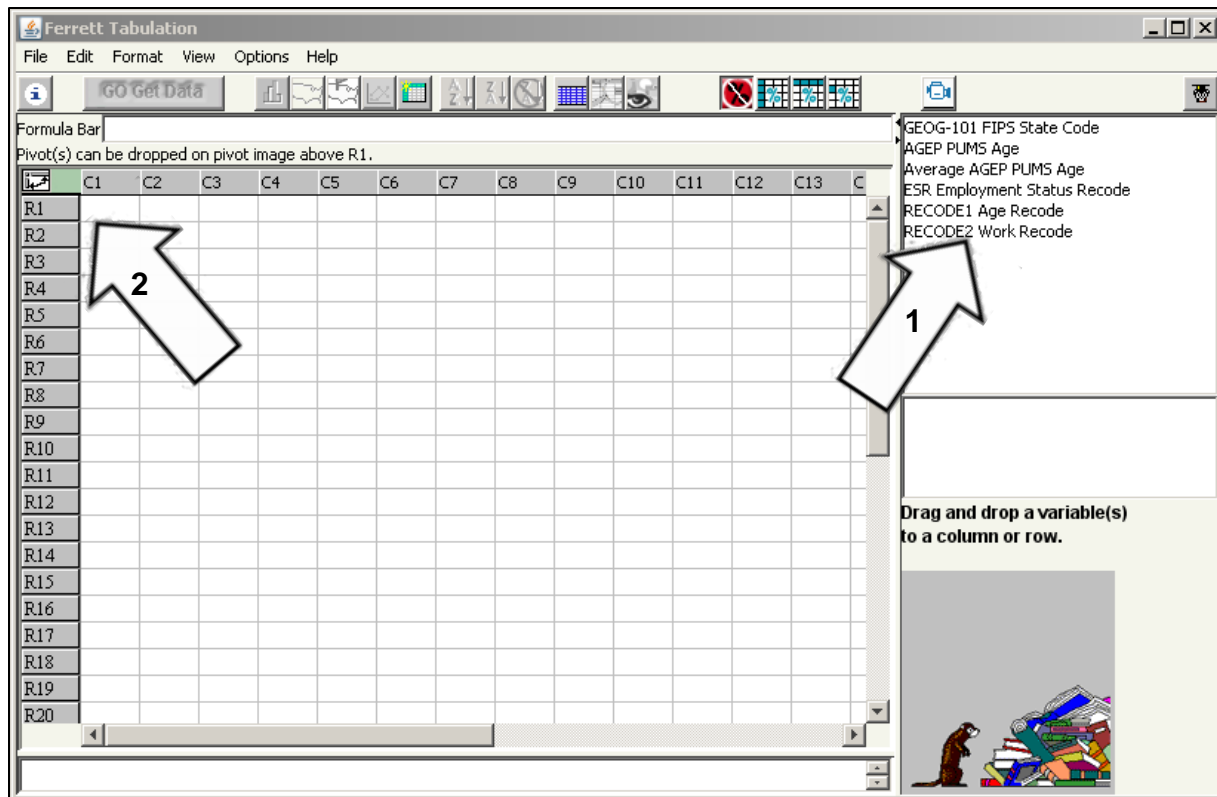


Figure 4-5: Ferrett Tabulation Screen

DataFerrett issues the following warning if the first variable is placed in a spreadsheet cell other than R1, C1.

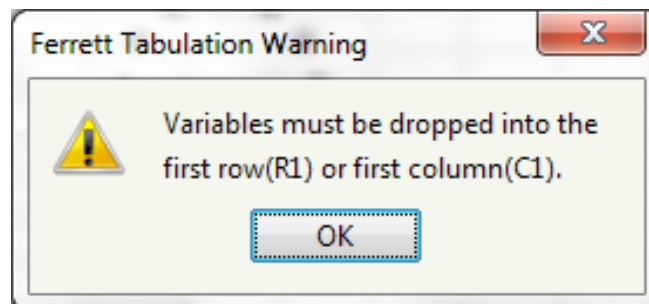


Figure 4-6: Ferrett Warning Message



HINT: To change the variable names in a table, double click the cell that contains the variable name and edit the text.

The following examples demonstrate how you can create a variety of table layouts. The first demonstrates how to create a simple crosstab and the second builds on the first by incorporating a geography variable and an instance of a dataset.

Example 4-1: Create a Simple Crosstab

This example uses variables from the 2006 – 2009 American Community Survey (ACS) Public Use Microdata Sample (PUMS) files and requires two recodes to be created based on the AGE and ESR variables.

Example 4-1: Create a Simple Crosstab		
Steps	What to do	Why do it?
1	Open the American Community Survey folder, select the year 2009, and highlight “View Variables.” Click “Select All Topics” then “Search.” You will be working with the Variables shown in Figure 4-7. To simplify this process, a Ferrett Session File that includes the two recode variables is available.	<p>All datasets needed for this example are contained in the American Community Survey. In addition to looking at labor status across age groups, you need to obtain geography variables for comparison.</p> <p>NOTE: Clicking on Geography opens a warning message that then opens the “Ferrett Geography Codebook.”</p>

Example 4-1: Create a Simple Crosstab		
Step	What to do	Why do it?
2.1 2.2	Go to the “Step 2” screen and click the “Make a Table” button (see Figure 4-1). When the spreadsheet appears, go to the “Options” Menu (2.1) and deselect the “Automatic Totals” option (2.2).	Unless “Automatic Totals” is turned off, a row labeled “Total” will automatically be added when the variable is dragged into position.
3	Drag the variable RECODE1 to cell R1C2, which results in the table shown in Figure 4-8.	The first column must be available for the categories for each variable.
4	Drag the variable RECODE2 to cell R1C1.	This will give you the desired layout shown in Figure 4-10.
5	Click on the “Go Get Data” button.	This action will populate the table with numbers (Figure 4-11).

NOTE: The same layout is obtainable by working in reverse order: drop RECODE2 in R1C1 first and RECODE1 in R1C2 next. You could have dropped variable RECODE2 in any row for column C1. Dropping the variable in any row below R2 in column C1 would have resulted in a layout with one or more empty rows above the data rows.

Step 1

Name	Variable Label	Availability
AGEP	PUMS Age	2006 - current
ESR	Employment Status Recode	2006 - current
RECODE1	Age Recode	2006 - current
RECODE2	Work Recode	2006 - current
PWGTP	PUMS person weight	2006 - current
GEOG-...	FIPS State Code	2006 - current

Figure 4-7: ACS DataBasket

Step 2

Figure 4-8: Turn off “Automatic Totals” in Options Menu

Step 3

	C1	C2	C3	C4
R1		Ages 0 to 17	Ages 18 to 64	Ages 65 and older
R2				
R3				

Figure 4-9: Results of Placing RECODE1 in Cell R1C2

Step 4

	C1	C2	C3	C4
R1		Ages 0 to 17	Ages 18 to 64	Ages 65 and older
R2	Not in the labor force		?	?
R3	In the labor force		?	?

Figure 4-10: Placing Second Recode Variable

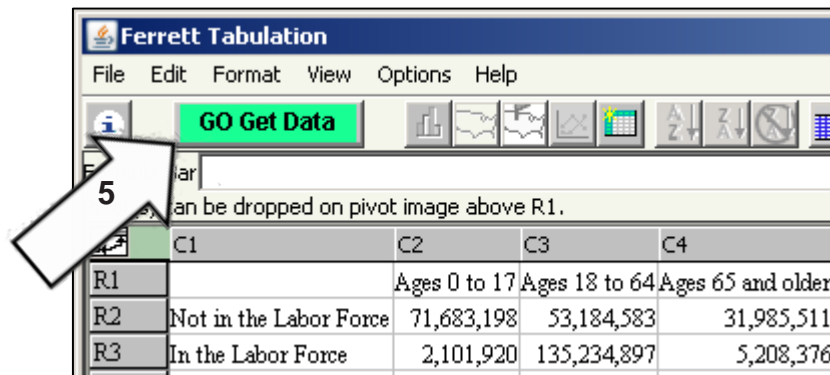
Step 5

Figure 4-11: Select “Go Get Data” to Populate Table

Example 4-2: Create 4-Dimensional Table Using State Geographies

Building on Example 4-1, the (Employment Status Recode) ESR variable has been added to the DataBasket from the ACS PUMS dataset. To create a 4-dimensional table, we will be adding instances for the ESR into the table. In order to accomplish this, you must return to the “Step 1” window and select the instances within the dataset. To do this, highlight the first and last instance while holding down the SHIFT key, as shown in Figure 4-12.

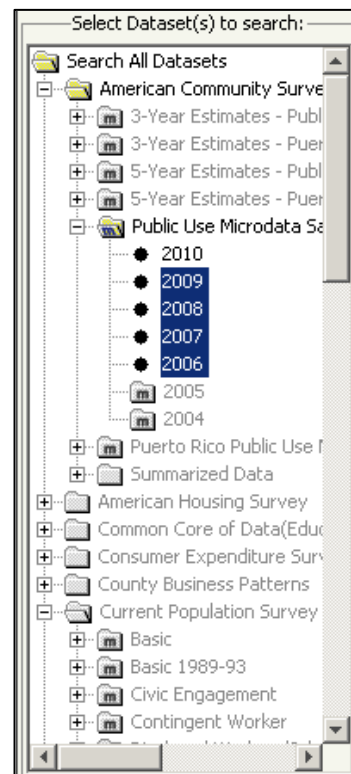


Figure 4-12: Selecting Instances

The continuation of Example 4-2 demonstrates how geography and an instance of a dataset are incorporated into a table.

Example 4-2: Create 4-Dimensional Table Using State Geographies		
Step	What to do	Why do it?
1.1 1.2	Drag RECODE1 (from the previous method, see arrow 1.1) to cell R1C2 (1.2).	The work recode, RECODE2, is “nested” inside RECODE1, the age dimension.
2.1 2.2	Select RECODE2 (2.1) and drag it to any of the columns in Row R1 (2.2). Doing this will result in the table layout shown in Figure 4-13.	Prior to doing this step, make sure “Automatic Cross Variable Explosion” is selected. Selecting “Automatic Cross Variable Explosion” allows you to see information on all instances. You can turn this feature on and off by accessing the Options menu (discussed later in this chapter).
3.1 3.2	To add the geography dimension, select the geography variable (“GEOG”) (3.1) and drag it into position R1C1 (3.2) yielding a new layout shown in Figure 4-14. Make sure to select the “Options” menu and turn off “Automatic Totals” (shown later in this chapter).	This layout effectively organizes the table layout to view the information by state, age and labor status.
4.1 4.2	To create the final layout, select the “Instances” variable (4.1) then drag it to any row in column C1 with a state name (4.2). Doing so will cause the warning message in Figure 4-16 to pop up. Also, make sure that “Automatic Cross Variable Explosion” is turned back on.	By clicking “yes” to the pop-up message, the final layout will be created (this step can take up to a few minutes to complete, see Figure 4-17 for the results). “Automatic Cross Variable Explosion” must be turned on so that the instances will be listed under ALL the states.
5	Once you have question marks (“?”) in the columns and rows of the table, click on the icon in the tool bar at the top of the Ferrett Tabulation window to retrieve the values for the variables in the table, or select File >Get Data .	This step is necessary to display the totals in the spreadsheet.

Steps 1 & 2

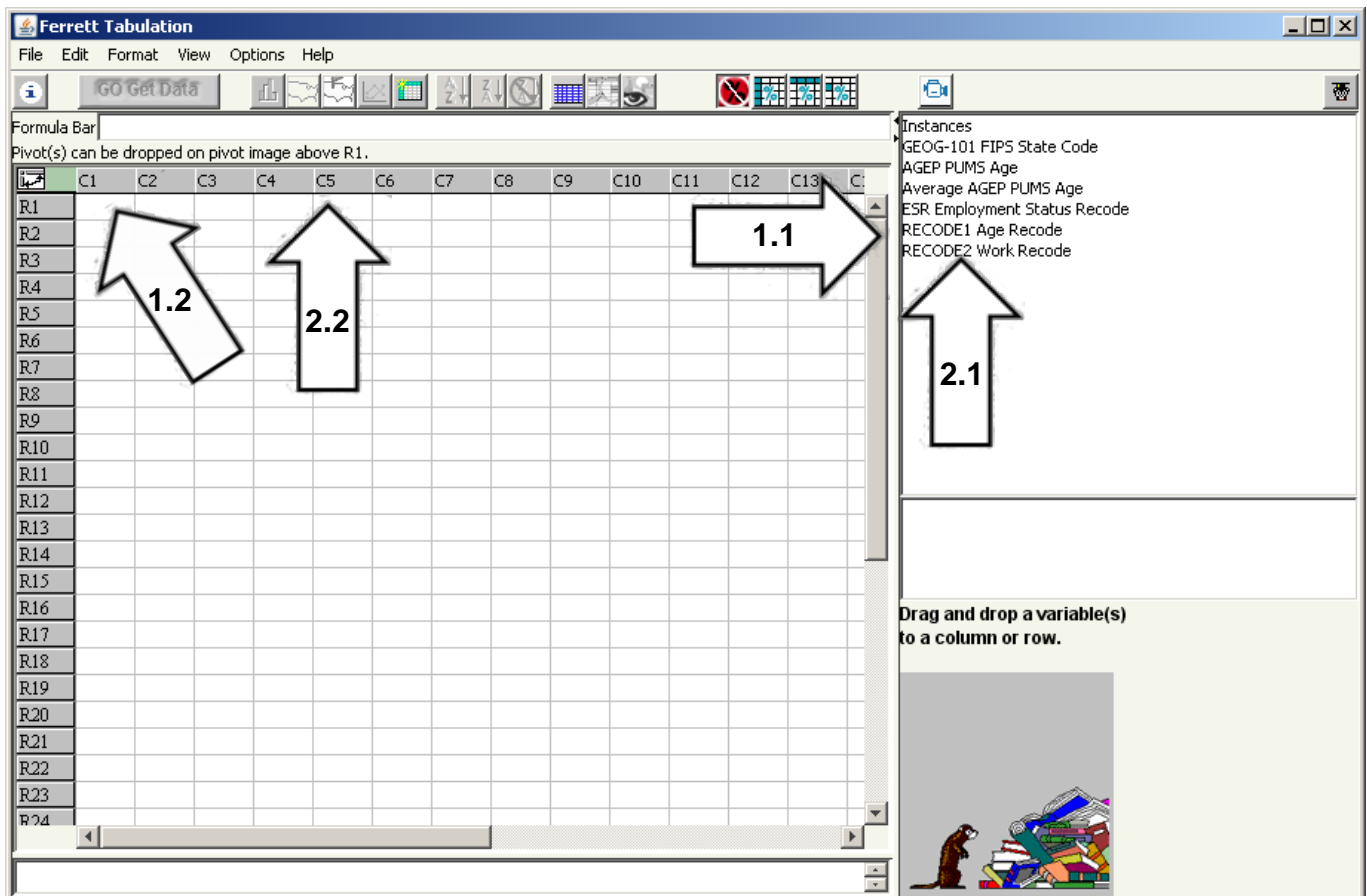


Figure 4-13: Nested Table Layout

Step 3

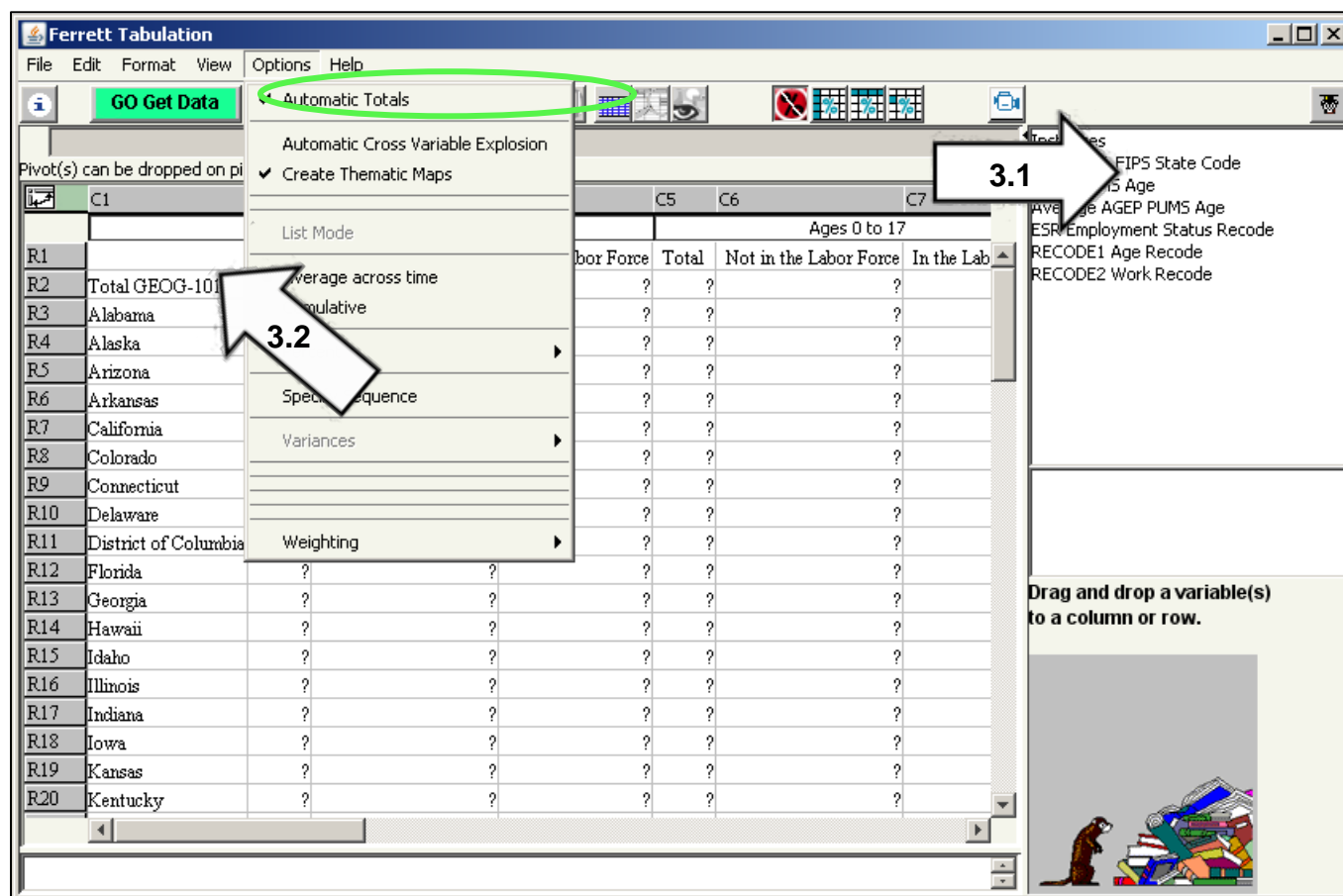


Figure 4-14: Populating Table with Data

Step 4

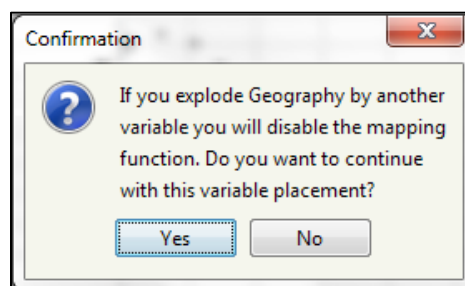


Figure 4-15: Pop-up Message Prior to Creating Final Layout

Step 4 (cont.)

Pivot(s) can be dropped on pivot image above R1.

4.1

	C1	C2	C3	C4	C5	C6
R1		Total RECODE1			Ages 0 to 1	
		Total	Not in the Labor Force	In the Labor Force	Total	Not in the Labor Force
R2	Total GEOG-101	?	?	?	?	?
R3	2009	?	?	?	?	?
R4	2008	?	?	?	?	?
R5	2007	?	?	?	?	?
R6	2006	?	?	?	?	?
R7	Alabama	?	?	?	?	?
R8	2009	?	?	?	?	?
R9	2008	?	?	?	?	?
R10	2007	?	?	?	?	?
R11	2006	?	?	?	?	?
R12	Alaska	?	?	?	?	?
R13	2009	?	?	?	?	?
R14	2008	?	?	?	?	?
R15	2007	?	?	?	?	?
R16	2006	?	?	?	?	?
R17	Arizona	?	?	?	?	?
R18	2009	?	?	?	?	?
R19	2008	?	?	?	?	?
R20	2007	?	?	?	?	?

4.2

Instances

- GEOG-101 FIPS State Code
- AGEP PUMS Age
- Average AGE PUMS Age
- ESR Employment Status Recode
- RECODE1 Age Recode
- RECODE2 Work Recode

Drag and drop a variable(s) to a column or row.


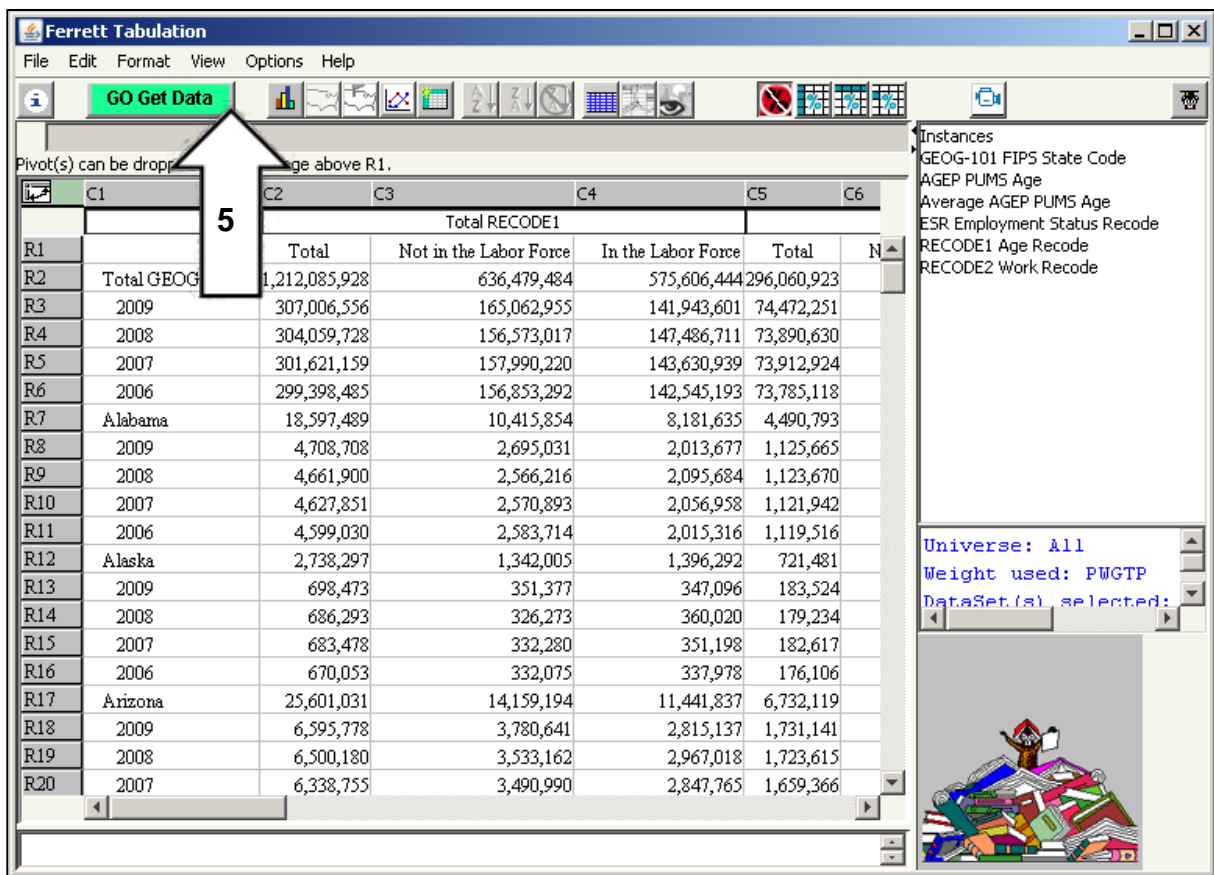


Figure 4-16: 4-Dimensional Table Layout

Step 5



Pivot(s) can be dropped here above R1.

	C1	C2	C3	C4	C5	C6
R1		Total	Not in the Labor Force	In the Labor Force	Total	N
R2	Total GEOG	1,212,085,928	636,479,484	575,606,444	296,060,923	
R3	2009	307,006,556	165,062,955	141,943,601	74,472,251	
R4	2008	304,059,728	156,573,017	147,486,711	73,890,630	
R5	2007	301,621,159	157,990,220	143,630,939	73,912,924	
R6	2006	299,398,485	156,853,292	142,545,193	73,785,118	
R7	Alabama	18,597,489	10,415,854	8,181,635	4,490,793	
R8	2009	4,708,708	2,695,031	2,013,677	1,125,665	
R9	2008	4,661,900	2,566,216	2,095,684	1,123,670	
R10	2007	4,627,851	2,570,893	2,056,958	1,121,942	
R11	2006	4,599,030	2,583,714	2,015,316	1,119,516	
R12	Alaska	2,738,297	1,342,005	1,396,292	721,481	
R13	2009	698,473	351,377	347,096	183,524	
R14	2008	686,293	326,273	360,020	179,234	
R15	2007	683,478	332,280	351,198	182,617	
R16	2006	670,053	332,075	337,978	176,106	
R17	Arizona	25,601,031	14,159,194	11,441,837	6,732,119	
R18	2009	6,595,778	3,780,641	2,815,137	1,731,141	
R19	2008	6,500,180	3,533,162	2,967,018	1,723,615	
R20	2007	6,338,755	3,490,990	2,847,765	1,659,366	

Instances
 GEOG-101 FIPS State Code
 AGE PUMS Age
 Average AGE PUMS Age
 ESR Employment Status Recode
 RECODE1 Age Recode
 RECODE2 Work Recode

Universe: All
 Weight used: PWGTP
 DataSet(s) selected:

Figure 4-17: 4-Dimensional Table Totals

NOTE: Other types of layouts can be created in DataFerrett in addition to the ones shown in the above examples. These alternatives include:

- Side-by-side layouts with no, or partial, cross tabulation
- Layouts that incorporate a “pivot” variable

Ferrett Tabulation Menus

File Menu



1. **New:** Closes the current table and opens a new empty spreadsheet while preserving the variable selection.
2. **Open:** Opens a previously saved table.
3. **Open in New Window:** Opens a new table in a new window.
4. **Save:** Saves the current table layout and DataBasket. As saving work in the open window will automatically overwrite an existing Ferrett tabulation file (.ftf) if opened in the current session, you are advised to use caution when saving a table session. The file is saved with the extension .ftf in TheDataWeb folder that is installed onto the C: drive when DataFerrett is first run.
5. **Save As:** Saves your current table layout and DataBasket as a Ferrett Tabulation File (.ftf), or it will save the current table and corresponding data (only) as an HTML file or text document (tab delimited or comma delimited).
6. **Get Data:** Retrieves the data for the variables that define the columns and rows in the table layout. This replaces the “?” in the data cells with values.
7. **Debug:** This feature is reserved for use by DataFerrett programmers.
8. **Print:** Prints a hard copy of the current spreadsheet.
9. **Print Selected:** Prints a hard copy of highlighted columns, rows, or cells.
10. **Table Properties:** Provides a space within the FTF created by the user in which to save additional documentation about the table being created. In addition, the entries in the properties Titles, Headers, and Footers will be stored in the PDF file created by the Generate PDF feature under the File menu. **NOTE:** This information will be used for any other table created unless the information is changed or cleared.
11. **Generate PDF:** Allows you to save the current table (table shell or populated table) as a .pdf file. The formatting used for the table in the pdf file makes the image of the table ready for later inclusion in documents created outside of DataFerrett.
12. **Graph:** Produces a graph for any set of previously highlighted data cells.

13. **Map:** Produces a thematic map for the highlighted cells at the selected sub-national geographic level, if a mappable dataset (either an aggregate dataset or microdata that has been summed to a mappable geographic level) has been laid out in the table.
14. **Map Points:** This option is not available to users at this time.
15. **TimeSeries Graph:** This option is not available to users at this time.
16. **Rename:** This option is not available to users at this time.
17. **Exit:** Closes the Ferrett Tabulation window.

Edit Menu



1. **Undo Dropping:** Removes the last variable dropped into the table.
2. **Cut:** Removes highlighted cells from the Ferrett Tabulation window and stores them in the paste buffer.
3. **Copy:** Copies highlighted content and stores it in the paste buffer.
4. **Paste:** Pastes highlighted content into another application outside of DataFerrett.
5. **Select All:** Selects all of the columns and rows in the spreadsheet.
6. **Hide:** Hides a selected row(s) or column(s).
7. **Unhide:** Restores all previously hidden rows and columns (see Figure 4-18).
8. **Clear:** Displays the slide-off menu in the Figure 4-19. The “Clear” slide- off menu options and their functionalities are:
 - a. Entire spreadsheet: All columns, all rows, all pivots, and all formulas defined by you are removed. The text of the query, shown in Figure 4-20, is also cleared.
 - b. All columns: The definitions of all columns starting with C2 are cleared. The definition of column C1 remains intact (i.e. the row definitions remain intact).
 - c. All rows: The definitions of all rows starting with R2 are cleared. The definition of row R1 remains intact (i.e. the column headers remain intact).
 - d. All pivots: The definitions of all pivots used are cleared.
9. **Delete:** Deletes selected row or column from the spreadsheet while adjusting the row or column numbering.
10. **Insert:** Inserts a blank row or column.
11. **Sort:** Sorts the highlighted row or column in ascending or descending order (Figure 4-21). When the column is deselected (highlighted) after a sort, it can be returned to its original order (see Figures 4-21 and 4-22).
12. **Column Spanners:** Displays the “Column Spanner Editing Dialog” pop-up window shown in Fig 4-20. There

are three required steps to create the desired column spanner:

- a.** Click on the Insert button of the dialog box (arrow 1 in Figure 4-23 below).
- b.** A modified version of the pop-up dialog box appears (see Figure 4-24) with a row inserted containing black cells where you must enter information (arrow 2 in Figure 4-24). Going from left to right double click on the black area for a cell. In the “Column Spanner” cell, enter a label. In “From” enter the left most column for the column spanner, in “To” enter the right most column.
- c.** Click on the “OK” button (arrow 3 in Figure 4-24) to create the column spanner right above the columns starting with the column in “From” and ending with the column in “To.”

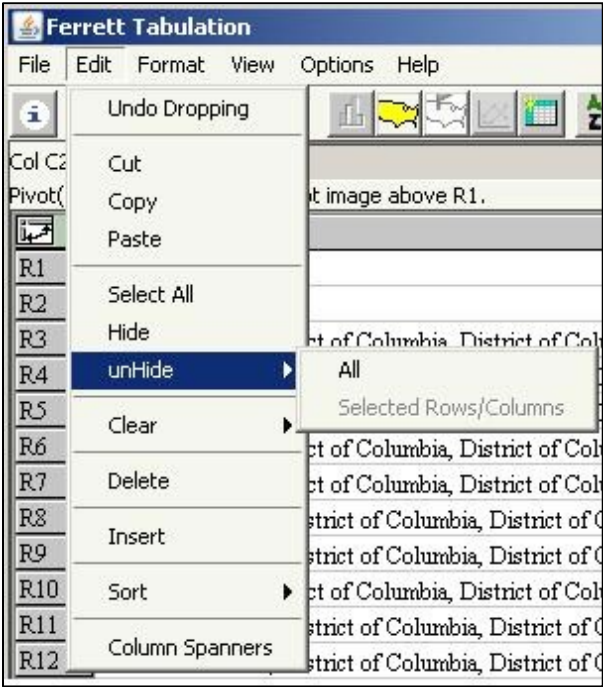


Figure 4-18: Unhide to Display All Rows

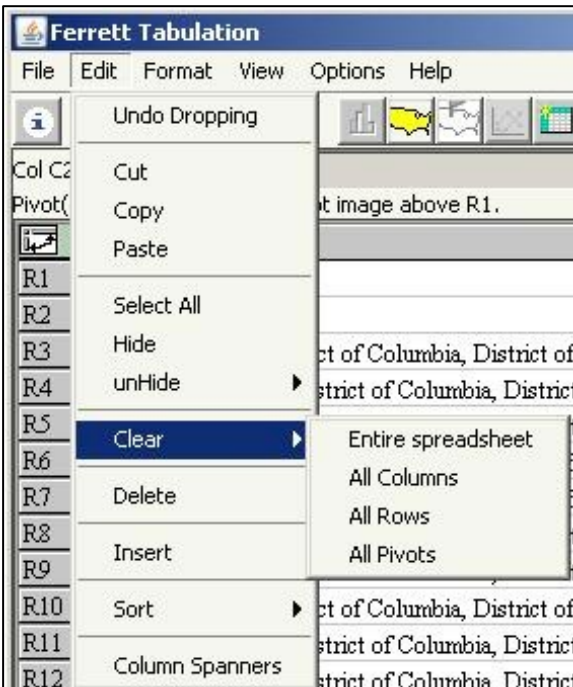


Figure 4-19: Clear Options



Figure 4-20: After Clearing Entire Spreadsheet

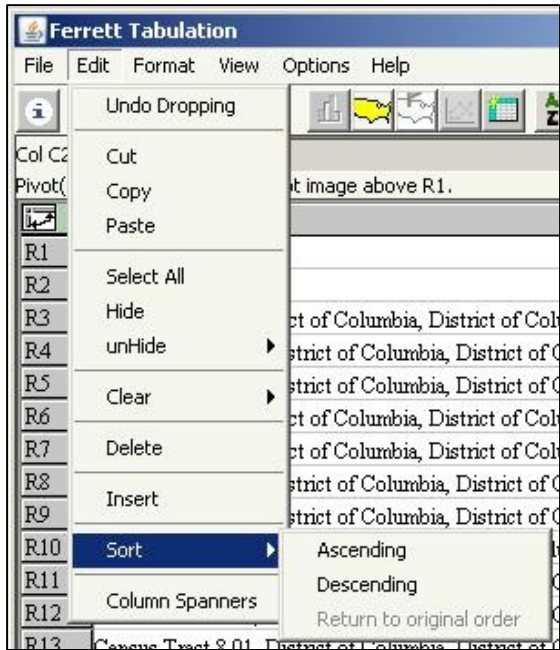


Figure 4-21: Sorting Rows in Table

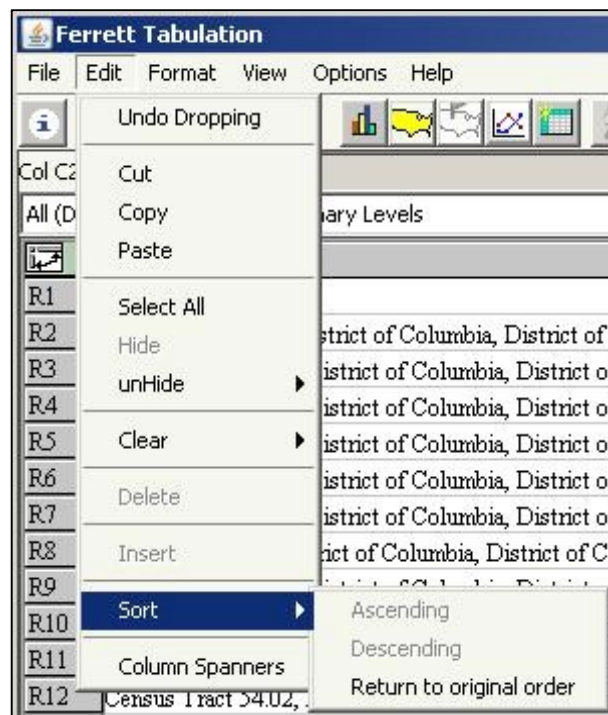


Figure 4-22: Return Columns to Original Order

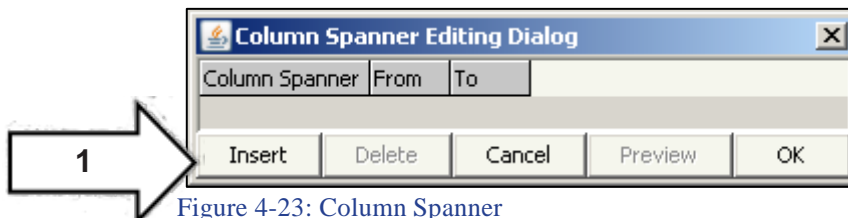


Figure 4-23: Column Spanner

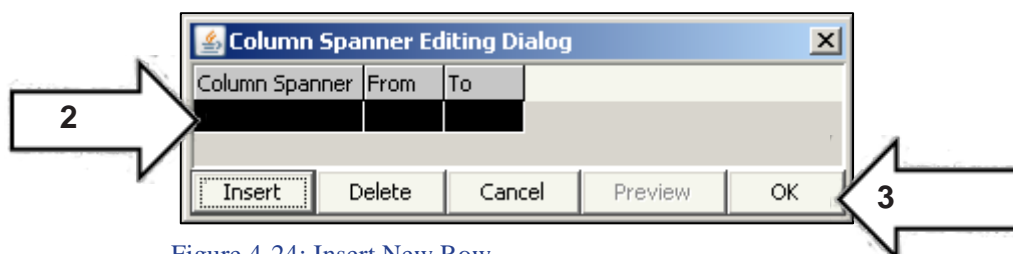
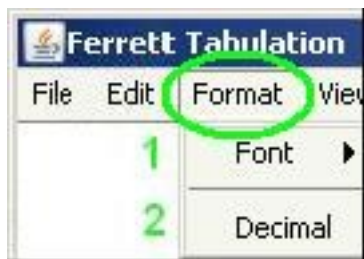


Figure 4-24: Insert New Row

Format Menu



1. **Font:** Displays a slide-off menu with three possible actions to modify the font of entries in a row or column: Type, Size or Style (Figure 4-25 displays the three options.)
2. **Decimal:** Allows formatting data cells into decimal point places from tenths to thousandths or as whole numbers from tens, hundreds, thousands or as code as shown in Figure 4-26. The “Number shown as code (no commas)” choice removes all commas from the display.

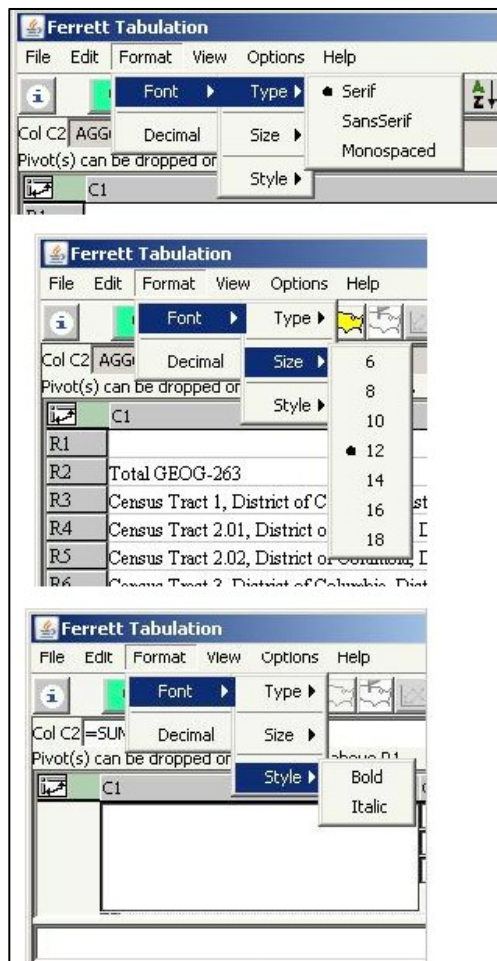


Figure 4-25: Format Font Menu

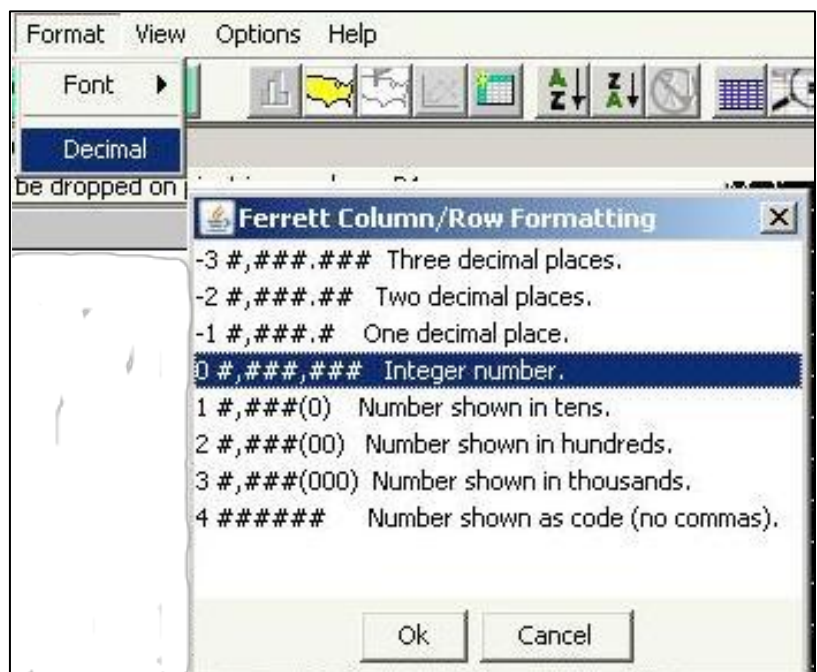


Figure 4-26: Format Decimal Cells

View Menu



1. **Tool Bar:** Acts as a toggle switch to show or hide the Tool Bar.
2. **Formula Bar:** Acts as a toggle switch for the Formula Bar in the table. If it is checked, the formula bar is visible and calculations can be performed. If it is not switched, the formula bar is hidden and no calculations can be performed on the table.
3. **Spreadsheet only:** If checked, only the spreadsheet will be displayed, hiding the variable list normally displayed at the far right of the table.
4. **Animation:** Displays three choices for sizing the Ferrett progress animation in the lower right hand portion of the screen (see Figure 4-27). It can be sized to normal, small or completely hidden ("hide"). This may be beneficial to users with smaller screen sizes or resolutions. (The "Hide" choice is not currently available.)

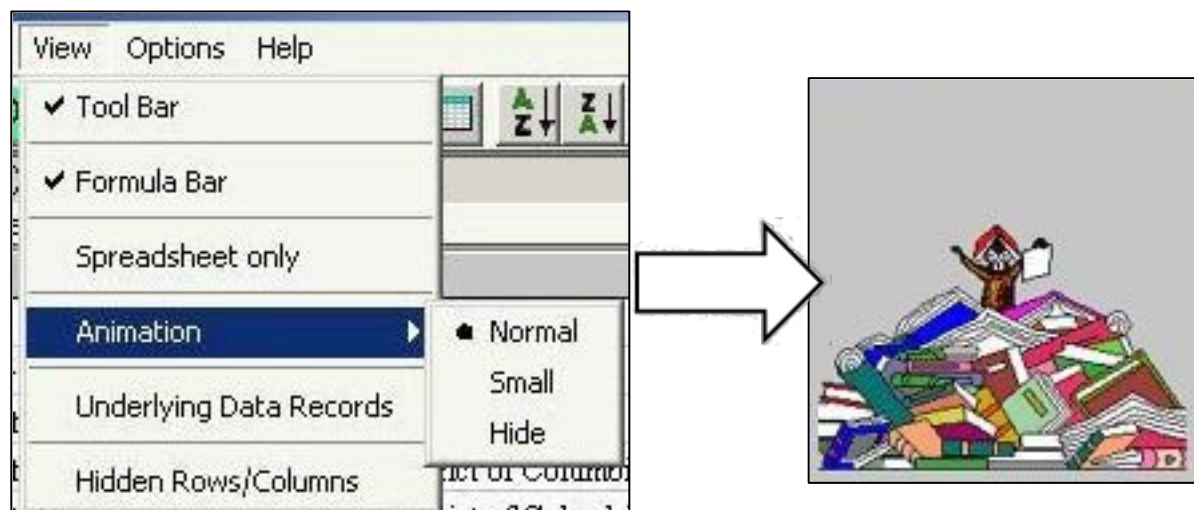


Figure 4-27: View Animation

5. **Underlying Data Records:** Opens a separate window displaying the microdata records that were tallied for the cell(s) selected. **NOTE:** This option is not available for variables that are not recode variables and have discrete values. The warning message shown in Figure 4-28 will appear if variables are chosen that do not meet these criteria.
6. **Hidden Rows/Columns:** Displays all hidden rows and columns with a white number on a gray background for the row number or column number. **NOTE:** This choice is different from the "Unhide" choice in the Edit menu that permanently removes the "hidden" status of these rows or columns.

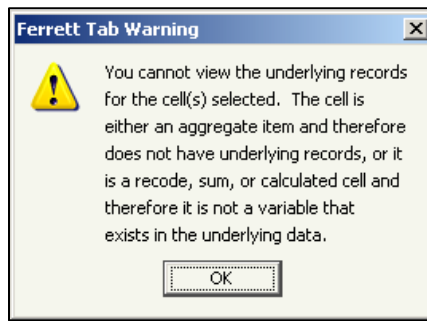
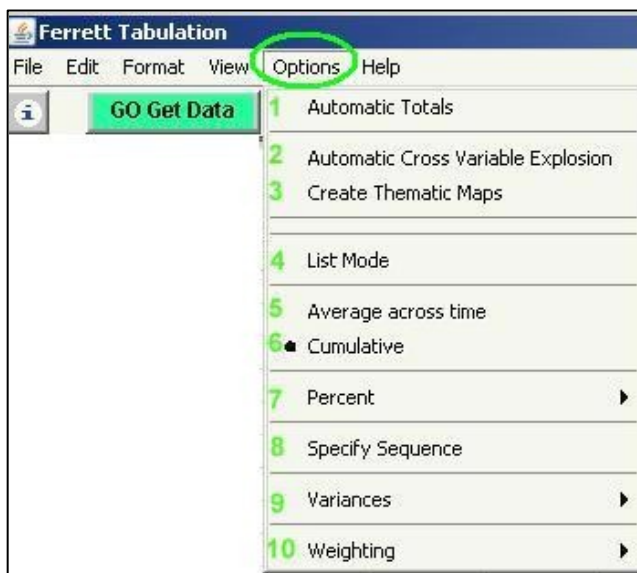


Figure 4-28: Underlying Records Message

Options Menu



1. **Automatic Totals:** Acts as a toggle to turn the feature on or off. If turned on, a row labeled "Total" will be added which will sum all the columns for the values as defined for the variable displayed along the top row of the table. This feature is activated only if checked before the table layout is defined (i.e., before the variables are dragged into position.)
2. **Automatic Cross Variable Explosion:** Acts as a toggle to turn the feature on or off. Automatic Cross Variable Explosion is only in effect if two or more variables are used in the table. If turned on and the first variable has been dragged to its position in the columns/rows and if the second variable is dragged to any column/row occupied by a value of the first variable, all values of the first variable will be partitioned into the values of the second variable (this is the "explosion.") This feature will have no effect if the second variable is dragged to an empty column/row.

3. **Create Thematic Maps:** Has the same effect as clicking on the U.S. map icon in the tool bar at the top of the window.
4. **List Mode:** Allows for the creation of a simple list of values for one or more variables side by side. Both the Automatic Totals and Automatic Cross Variable Explosion do not apply when List mode is activated. This feature is often used with aggregated datasets in conjunction with the Advanced SQL Option in the “Step 2” screen to create a list of geographic areas meeting certain criteria as specified by a user-defined query. List Mode can also be used with microdata datasets. **NOTE:** Column C1 may not be used when List mode is turned on.
5. **Average Across Time:** Displays the average across time for a variable(s) in a table where selected multiple instances are used to compute the average.
6. **Cumulative:** Displays the cumulative total across a time period for a variable(s) in a table where selected multiple instances are used to compute the cumulative total.
7. **Percent:** Allows percentage calculations in the same manner as the percentage options on the Ferrett Tabulation toolbar. Figure 4-29 below shows the slide-off menu of choices for Percent.

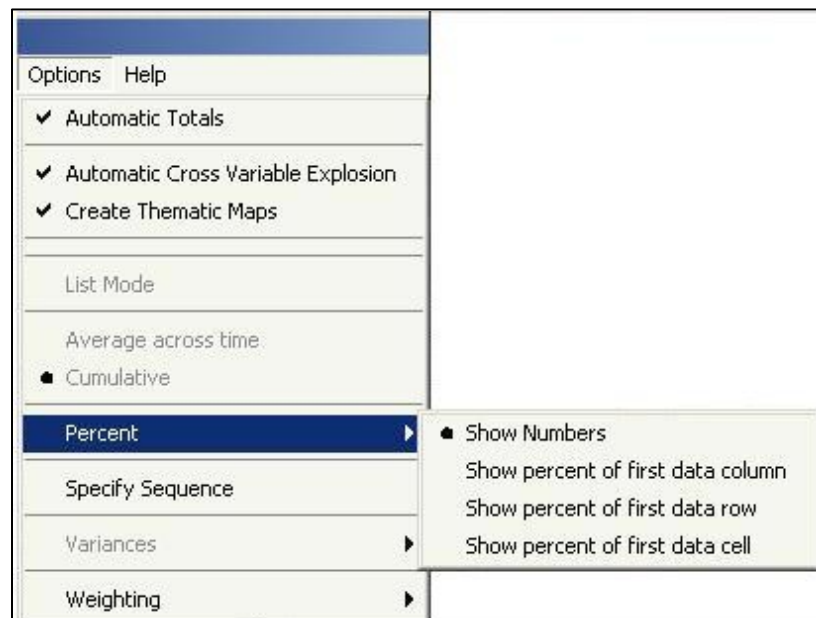


Figure 4-29: Percent Menu Choices

8. **Specify Sequence:** Enables control of the order of computations within the table.
9. **Clicking on Variances:** This feature is currently under testing.
10. **Weighting:** Clicking on or “mousing over” (see Figure 4-30) will display the choices for applying weights during the tabulation. Data that is pre-aggregated (e.g., Census 2000 Summary File 3 or the ACS

Summarized Data) has only one choice: unweighted because the weighting was already applied in computing these pre-aggregated data. If weights are available for a dataset (e.g., an ACS or CPS Public Use Microdata Sample (PUMS) dataset) and there is more than one type of weight variable, DataFerrett assigns a default weight for the table computation. If two variables with different default weights (i.e., person vs. household) are dropped into the table, you must choose one of the weights before the data can be retrieved. Once you have populated the table (i.e. clicked on “Go Get Data), the weight used for the tabulation is added to your DataBasket. Fig 4-30 shows an example of the weighting feature displaying multiple weight choices.

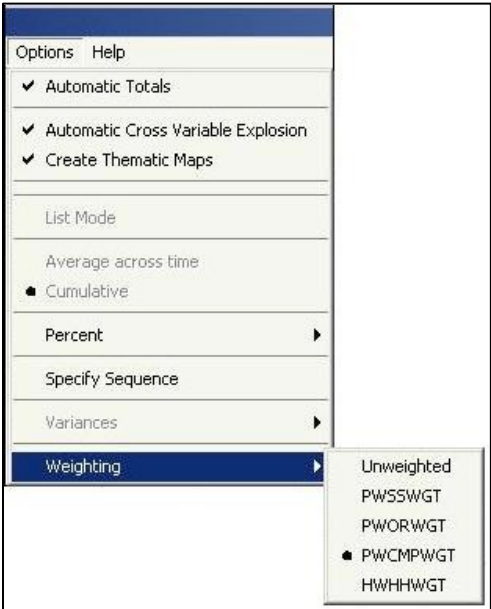
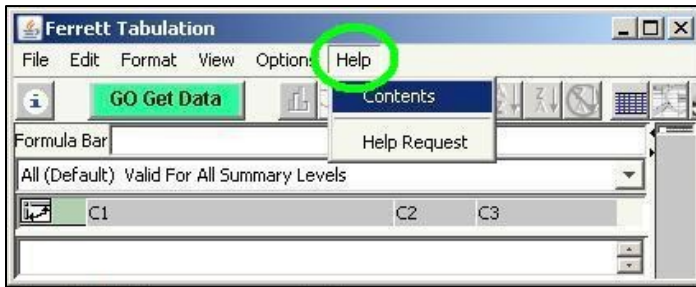


Figure 4-30: Weighting Options

Help Menu



1. **Contents:** Selecting this option brings you to the DataFerrett Spreadsheet Help page.
2. **Help Request:** Selecting this option opens a form to use when sending an email request for assistance (see Figure 4-31).

Figure 4-31: Help Request window

Creating New Columns Using Formulas

The Ferrett Tabulation window provides a formula area for columns and rows in a manner quite similar to spreadsheet applications such as Excel. However, there are currently no built-in functions available in Data- Ferrett, with the exception of the square root (SQRT*.) Computations can be created to add, subtract, multi- ply, divide, or calculate a square root of numbers. In Figure 4-33, the formula is dividing the contents of row R3 by those in R2 and multiplying the resulting ratio by 100.

Formulas refer to entire row or columns. Refer to rows as R1, R2, etc., and refer to columns as C1, C2, etc. Because of this, a formula is only valid when you click on a column or row heading, as shown in Figure 4-33 where you have clicked on column C12.

There are three types of formulas possible within DataFerrett. All formulas must begin with an “=” sign.

1. Computational

Computational formulas (=comp or =sum) include adding, subtracting, multiplying, dividing and deriving square roots. Simple conditions are allowed for columns or rows. Examples of computational formulas are:

```
=comp(R2+(R2*R3))
=comp((C2/C3)*100.0)
=sum(C2,C3,C4,C5)
=comp(sqrt(C3))
```

2. Ranking

Ranking displays the ordinal rank of each cell in a column (row) from smallest to largest. For example, if C4 is a data column, then defining the formula “=rank(C4)” in the formula bar of an empty column will display the relative rank of each cell in C4 in the column containing the rank() formula.

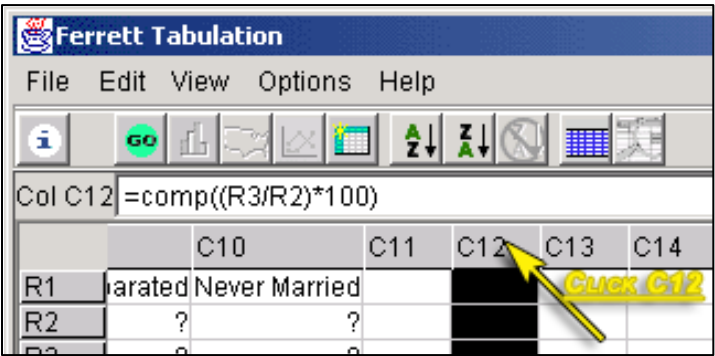


Figure 4-32: Create Column with a Formula

3. Simple Conditions

Examples:
=IF((C2 > 0),C3,C2)
=IF((C2>0) then C3 else
C2) : This feature is not
currently available.

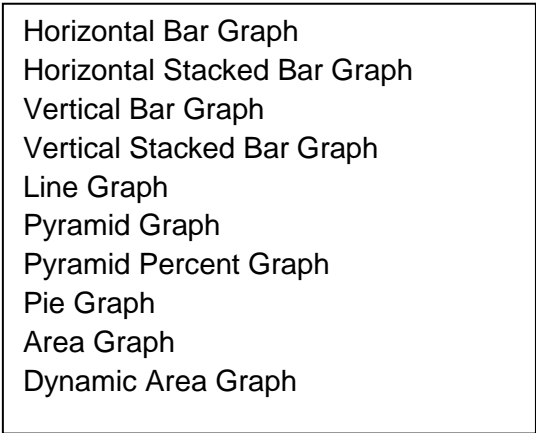
DataFerrett computations are capable of adding missing data, which are shown as [miss1]. It can add missing data for up to 5 instances: [miss1 - miss5] which have specific values and specific meanings.

You can save table definitions for later usage. Select File > Save As>: to save the file. The variables from the DataBasket and the layout of spreadsheet will be saved as an .ftf file (Ferrett Table File). The numbers will not be shown when you reopen the table, but the [GO] button will be active. You can also save and open the Table layout by clicking on the drop down menu under [File].

CHAPTER 5: MAKE A GRAPH

The “Step 2: DataBasket/Download/Make a Table” offers many options for displaying variables that have been selected and added to the DataBasket in “Step 1: Select Dataset & Variable.” These options are available via the Ferrett Tabulation Toolbar, accessed by choosing the “Make a Table” icon, and includes the ability to create a graph from a table. This chapter describes the types of graphs a user can make and details the different formatting and editing features that the user can employ to make a graph more readable.

The types of graphs available in DataFerrett are listed below. The type of graph that is selected is dependent upon the type of data that is displayed.



- Horizontal Bar Graph
- Horizontal Stacked Bar Graph
- Vertical Bar Graph
- Vertical Stacked Bar Graph
- Line Graph
- Pyramid Graph
- Pyramid Percent Graph
- Pie Graph
- Area Graph
- Dynamic Area Graph

Figure 5-1: Types of Graphs Available

Ferrett Graph Window

The type of graph is selected in the Ferrett Graph window, which is shown below.

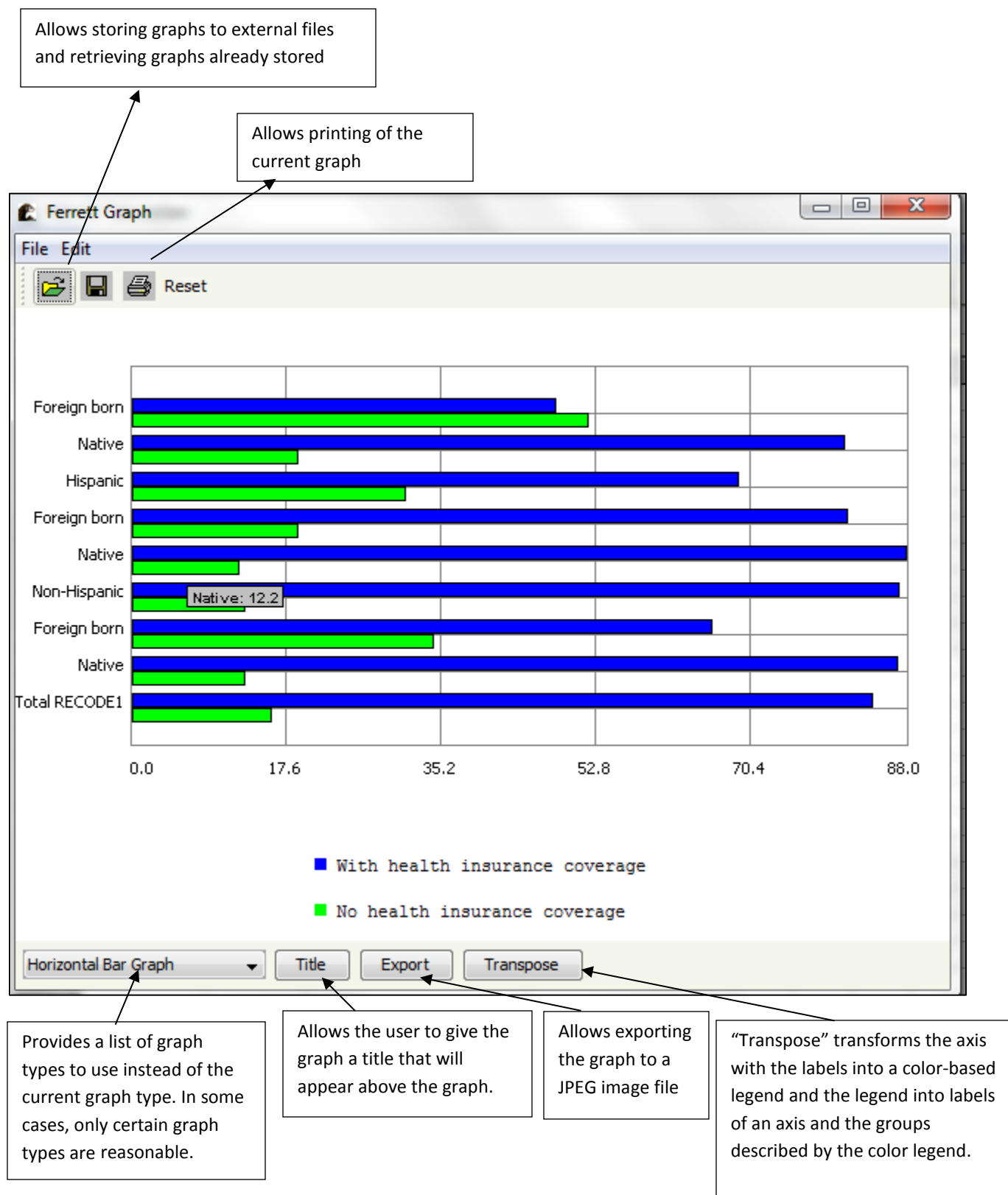
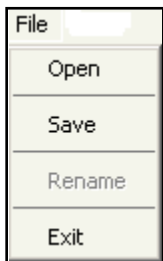


Figure 5-2: Ferrett Graph Window

File Menu



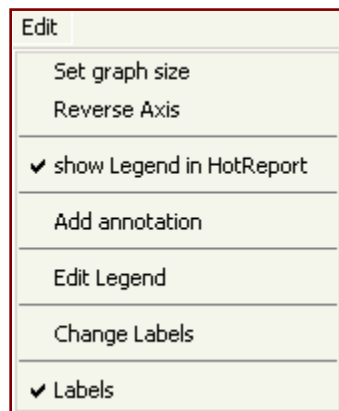
Open: Allows a saved graph file with the extension .graph to be imported.

Save: Allows a graph to be saved in the .jpg image file format.

Rename: This feature is currently unavailable.

Exit: Exits the graph window. **NOTE:** Exiting does NOT save a graph.

Edit Menu



Set graph size: Allows you to re-size the width and height of the graph window in a dialog box. The size is retained when the graph is printed or saved as an image file.

Reverse Axis: Reverses the order in which the characteristics are displayed on the X (horizontal) or Y (vertical) axis, depending on which axis is used for displaying the labels for the characteristics. In a horizontal bar chart, selecting “Reverse Axis” will reverse the order of the labels shown on the Y axis. In a vertical bar chart, selecting “Reverse Axis” will reverse the order of the labels shown on the X axis.

Show legend in HotReport: This feature is only available to DataFerrett users with permission to author DataFerrett Hot Reports.

Add annotation: Allows the creation and placement of text dialog box on a graph. Transposing the graph has no effect on the position of the annotation.

Edit legend: Allows you to change the definition of the colors used in the legend.

Change Labels: Allows you to change labels for the variable values in a dialog box.

Labels: Acts as a toggle switch to make labels appear or disappear on graph.

Example 5-1: Compare Health Insurance Coverage for Native vs. Foreign-born Population

Using the 2010 PUMS file from the American Community Survey as the dataset instance, this example studies health insurance coverage for the native born vs. the foreign-born population. In the following example, you are going to give special emphasis to the foreign-born Hispanic population.

Three variables are needed in the DataBasket for this example: NATIVITY; HISP (Hispanic origin); and HICOV, a built-in recode for health insurance coverage. We also need to recode the HISP variable into a variable with two values: Hispanic and Non-Hispanic. Figure 5-3 shows the DataBasket after the user has taken these actions.

Name	Variable Label	Availability
HICOV	Health insurance coverage recode	2008 - current
HISP	Hispanic recode	2006 - current
NATIVITY	Nativity	2006 - current
RECODE1	Hisp recode	2006 - current

Figure 5-3: DataBasket after HISP Recode Defined

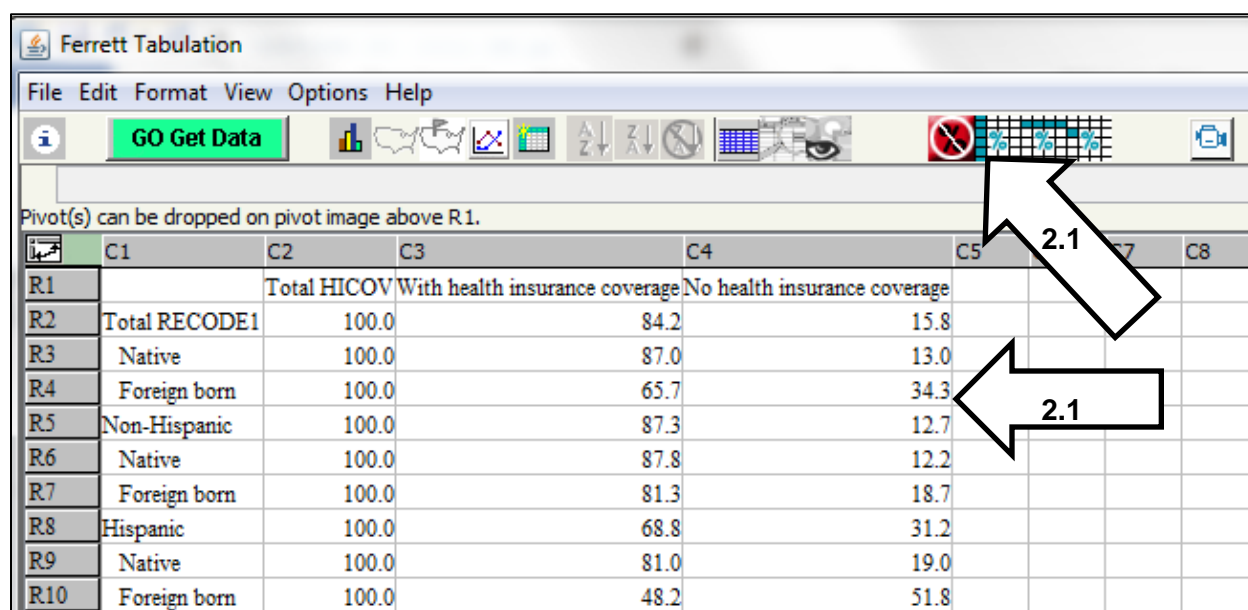
Example 5-1: Compare Health Insurance for Native vs Foreign-born		
Steps	What to do	Why do it?
1	Create the table shown in Figure 5-4. (Refer to Chapter 4: Make a Table for instructions on creating and populating a table.)	The creation of a table activates the graphing function when the Go Get Data button is clicked.
2.1 2.2	Convert the numbers in columns C3 and C4 to percents of column C2. This is done by clicking on the leftmost percent icon in the toolbar above the table. Columns C3 and C4 are now expressed as percentages of column C2.	It is difficult to determine the relative chance of having health insurance for each group described in the rows of the table when looking at the weighted counts for each group. By converting these counts to percentages, you have a clearer understanding of the information.
3.1 3.2	Highlight columns C3 and C4 and click on the graph icon in the toolbar.	By highlighting the columns, you are telling DataFerrett what information you want in your graph. In some cases, it may be more useful to highlight several columns for a single row in a table.
4	Place the mouse over the last bar in the graph (Hispanic foreign born with no health insurance) to see the numeric value.	This feature helps you analyze the magnitude of a difference between elements (e.g., lengths of bars) in a graph.

Step 1

	C1	C2	C3	C4
R1		Total HICOV	With health insurance coverage	No health insurance coverage
R2	Total RECODE1	309,349,689	260,556,127	48,793,562
R3	Native	269,432,814	234,339,569	35,093,245
R4	Foreign born	39,916,875	26,216,558	13,700,317
R5	Non-Hispanic	258,620,119	225,654,338	32,965,781
R6	Native	237,520,349	208,501,512	29,018,837
R7	Foreign born	21,099,770	17,152,826	3,946,944
R8	Hispanic	50,729,570	34,901,789	15,827,781
R9	Native	31,912,465	25,838,057	6,074,408
R10	Foreign born	18,817,105	9,063,732	9,753,373

Figure 5-4: Table Created with Variables in Figure 5-3

Step 2



Pivot(s) can be dropped on pivot image above R1.

	C1	C2	C3	C4	C5	C7	C8
R1		Total HICOV	With health insurance coverage	No health insurance coverage			
R2	Total RECODE1	100.0	84.2	15.8			
R3	Native	100.0	87.0	13.0			
R4	Foreign born	100.0	65.7	34.3			
R5	Non-Hispanic	100.0	87.3	12.7			
R6	Native	100.0	87.8	12.2			
R7	Foreign born	100.0	81.3	18.7			
R8	Hispanic	100.0	68.8	31.2			
R9	Native	100.0	81.0	19.0			
R10	Foreign born	100.0	48.2	51.8			

Figure 5-5: Values Expressed as Percentages

Step 3

The screenshot shows the Ferrett Tabulation software interface. At the top is a menu bar with 'File', 'Edit', 'Format', 'View', and 'Options'. Below the menu bar is a toolbar with various icons, including a 'GO Get Data' button. A callout arrow labeled '3.2' points to the 'GO Get Data' button. Below the toolbar, the text 'Col C3 HICOV=1' is displayed, followed by the instruction 'Pivot(s) can be dropped on pivot image above R1.' The main data table has columns C1 through C8 and rows R1 through R11. A callout arrow labeled '3.1' points to the data in row R5, column C4. The data table is as follows:

	C1	C2	C3	C4	C5	C6	C7	C8
R1		Total HICOV	With health insurance coverage	No health insurance coverage				
R2	Total RECODE1	100.0	84.2	15.8				
R3	Native	100.0	87.0	13.0				
R4	Foreign born	100.0	65.7	34.3				
R5	Non-Hispanic	100.0	87.3	12.7				
R6	Native	100.0	87.8	12.2				
R7	Foreign born	100.0	81.3	18.7				
R8	Hispanic	100.0	68.8	31.2				
R9	Native	100.0	81.0	19.0				
R10	Foreign born	100.0	48.2	51.8				
R11								

Figure 5-6: Selection of Data for Use in Graph

Step 3, cont.

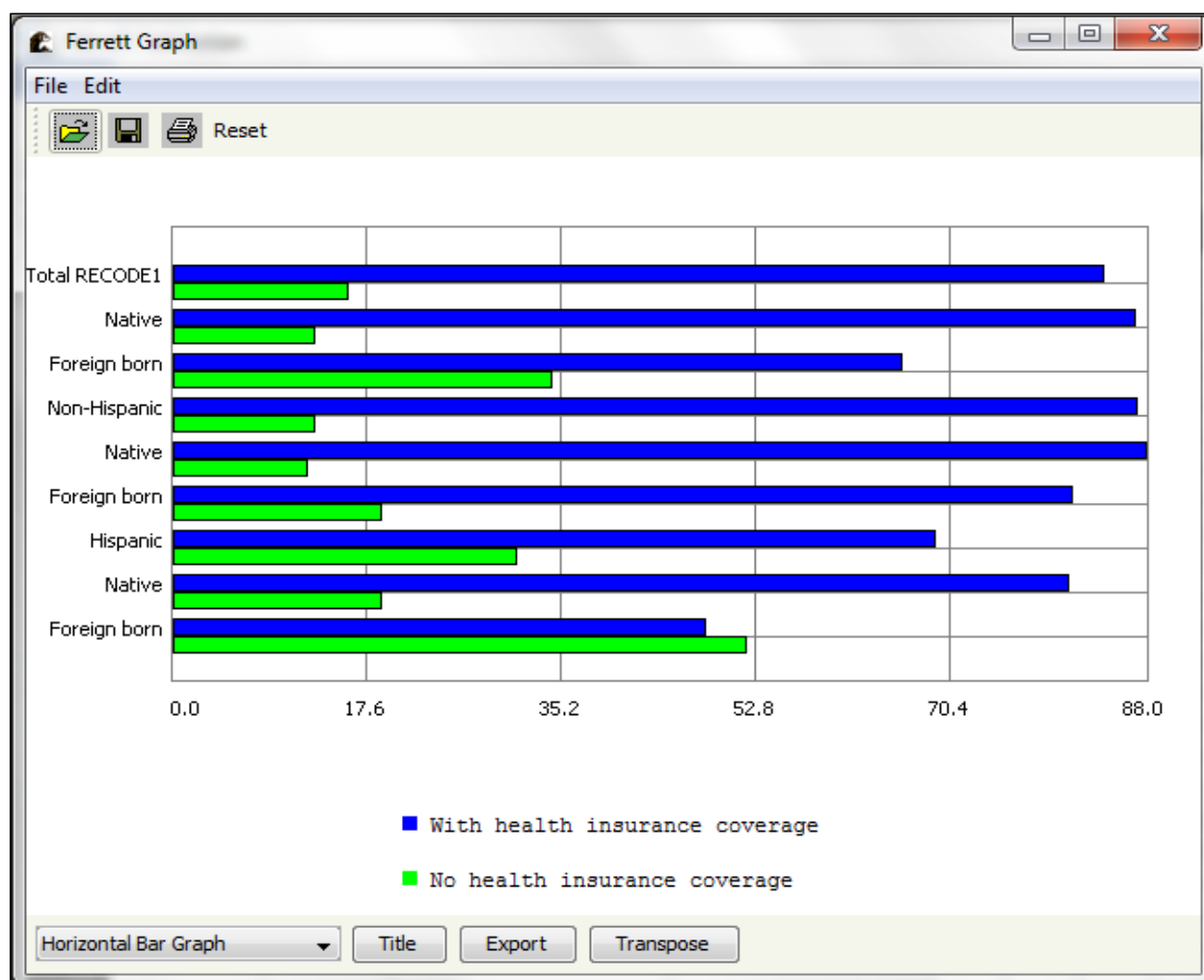


Figure 5-7: Initial Graph Displayed after Graph Icon Selected

Step 4

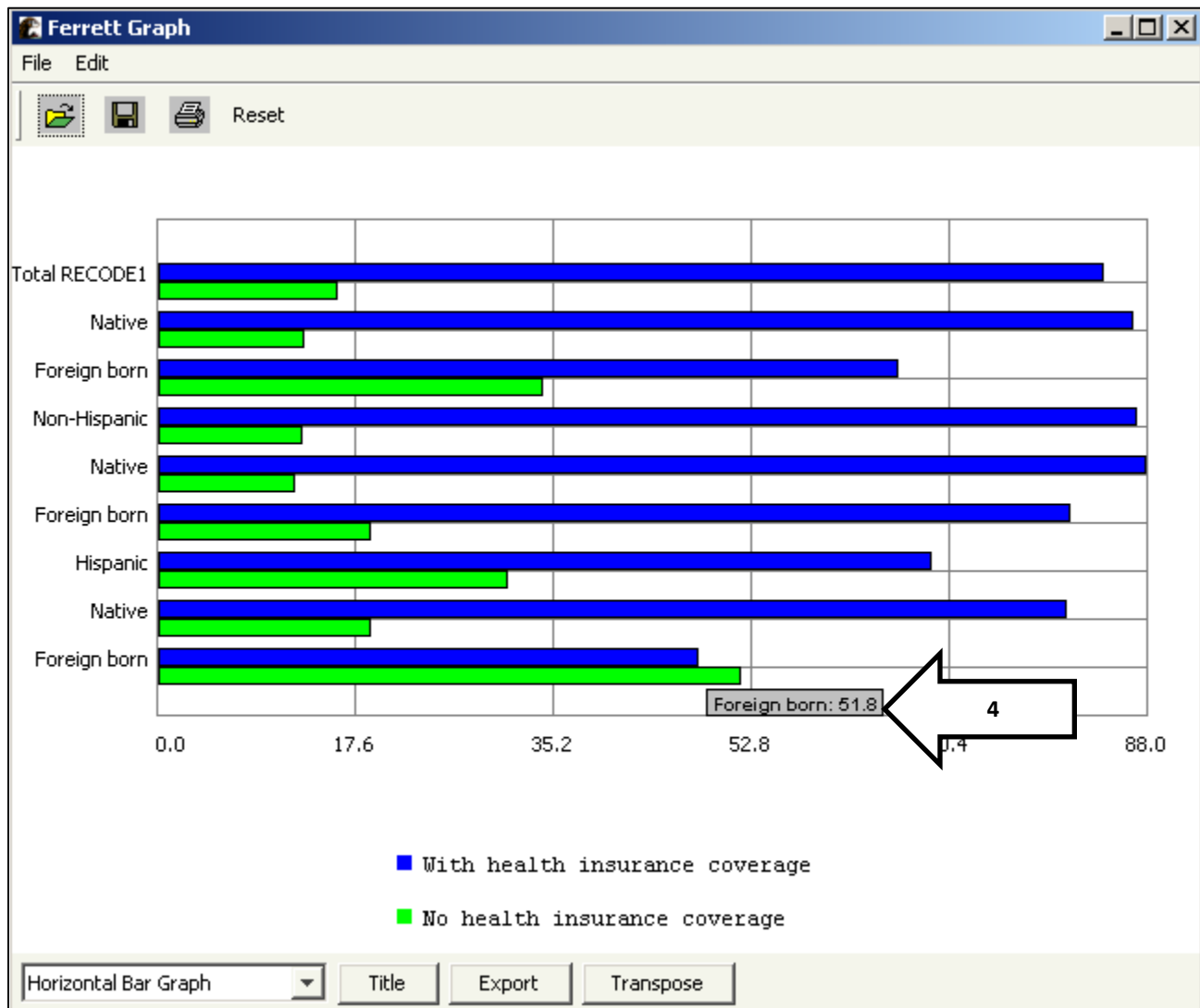


Figure 5-8: Exact Value of Graph Element Shown

Example 5-2: Using Ferrett Graph Features to Enhance Graph

The next example illustrates several of the Ferret Graph features that can be used to make a graph more readable, beginning with the graph shown in Figure 5-8. This graph did not have any features of the Ferrett Graph window applied to it in Example 5-1.

Example 5-2: Improve Appearance and Readability of Graph from Example 5-1		
Steps	What to do	Why do it?
1.1 1.2	Open the Edit menu and click on "Change Labels" (Figure 5-9). Figure 5-10 shows the current labels that are then changed in Figure 5-11. Figure 5-12 shows the graph with the	This enables clearer and unambiguous labels for each row.

	new labels.	
2	Click on Title and add the title “Pct with Health Insurance by Hispanic Origin and Nativity.”	Adding a title defines the graph.
3	Click on tab at bottom left to change graph type to “Horizontal Stacked Bar Graph.”	Modifying graph type often produces a clearer and more understandable representation of data.

Step 1

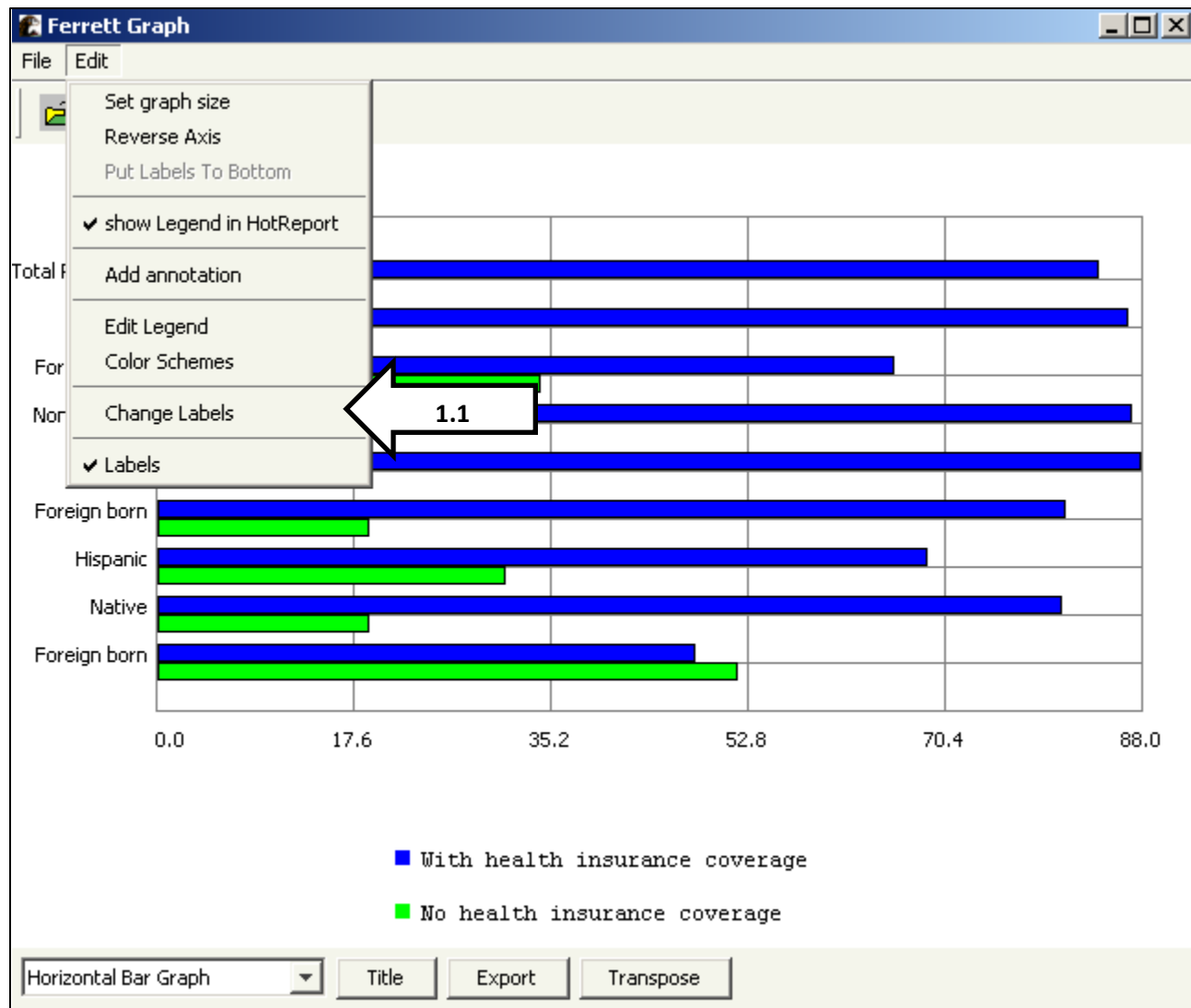


Figure 5-9: Ferrett Graph Edit Menu

Step 1, cont.

?

Total RECODE1

Native

Foreign born

Non-Hispanic

Native

Foreign born

Hispanic

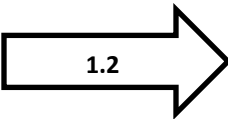
Native

Foreign born

OK

Cancel

Figure 5-10: Current Labels



?

Total Population

Native

Foreign born

Non-Hispanic

Non-Hispanic Native

Non-Hispanic Foreign born

Hispanic

Hispanic Native

Hispanic Foreign born

OK

Cancel

Figure 5-11: Modified Labels

Step 1, cont.

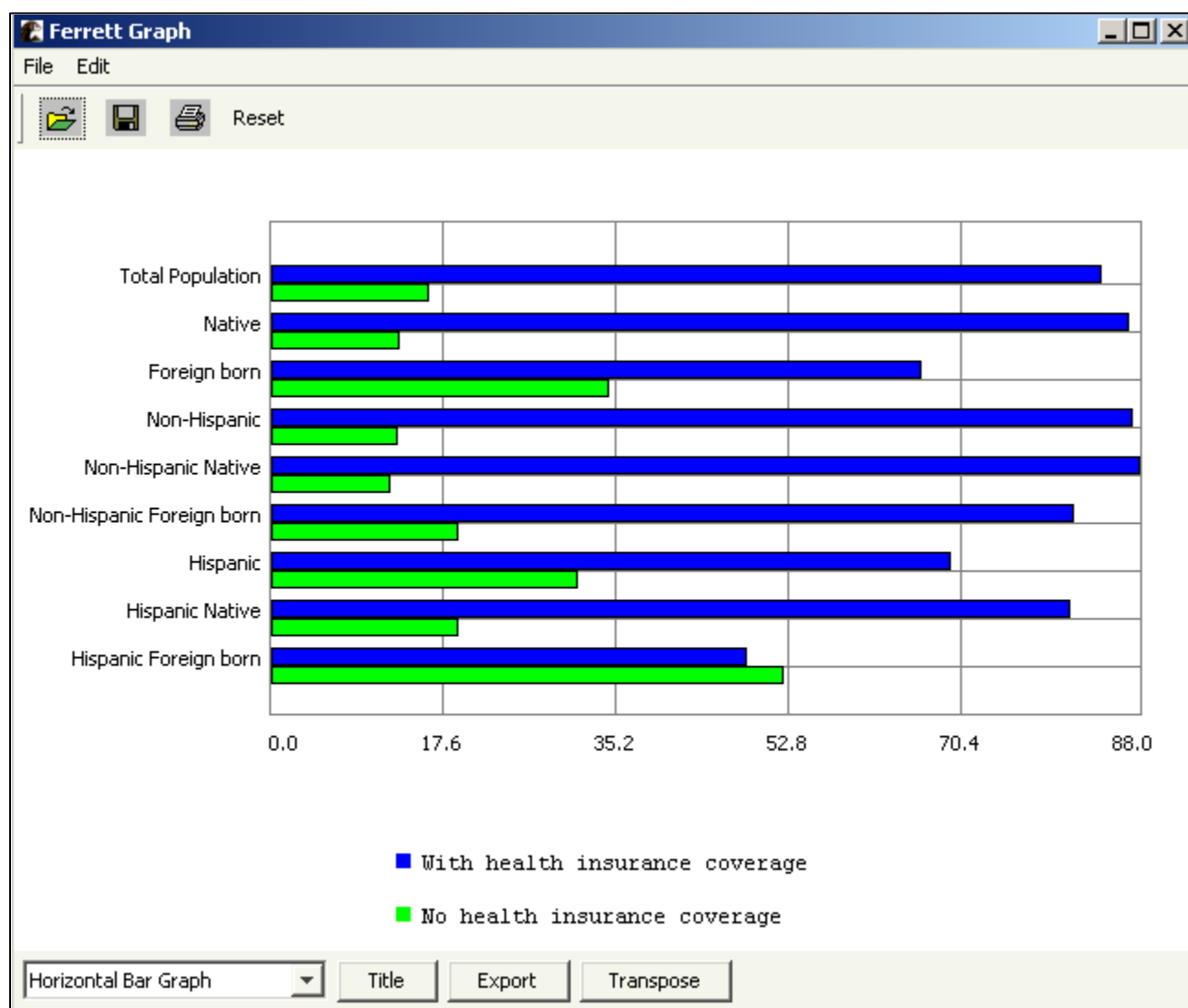


Figure 5-12: Graph Modified with New Labels

Step 2

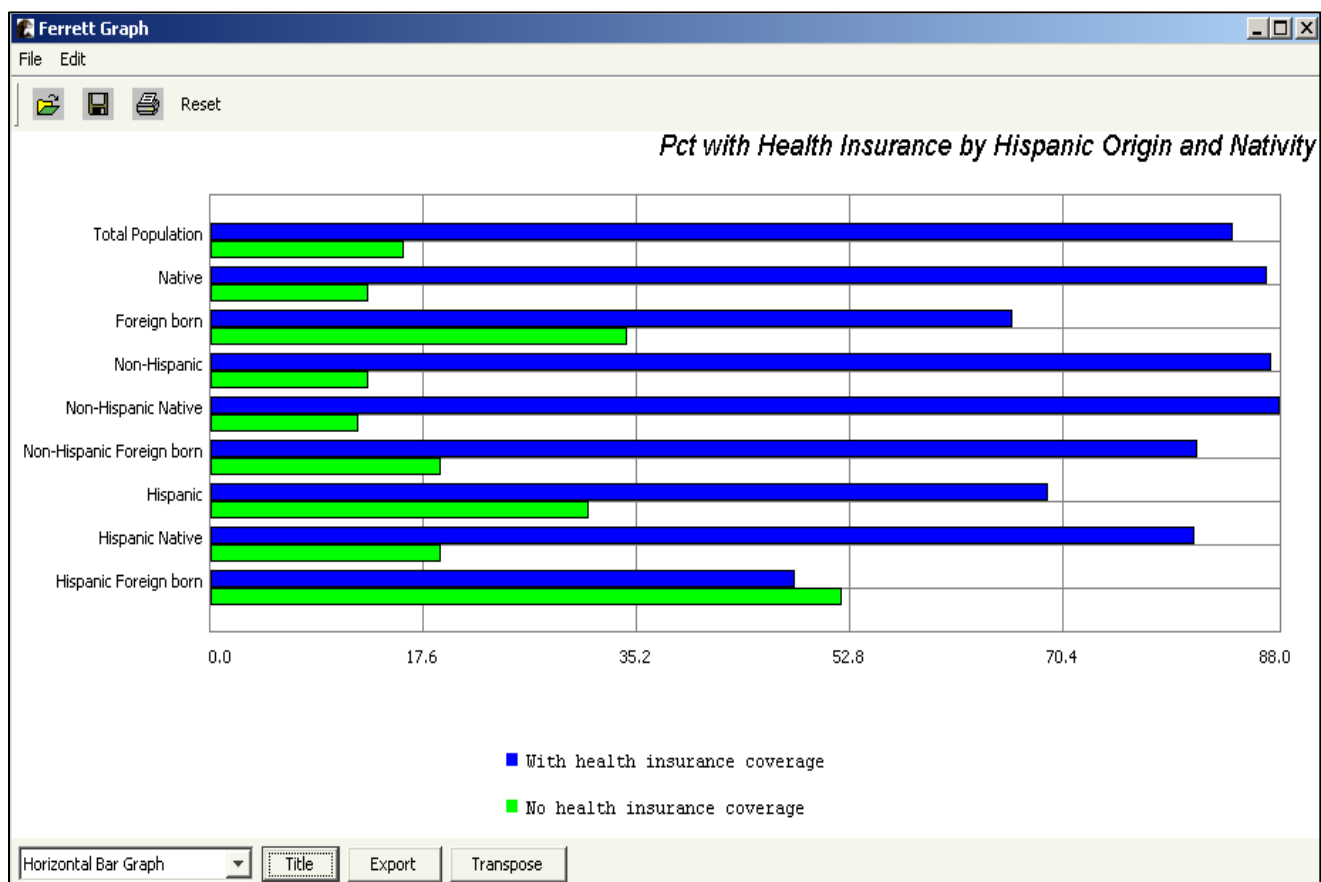


Figure 5-13: Modified Graph with Title Added

Step 3

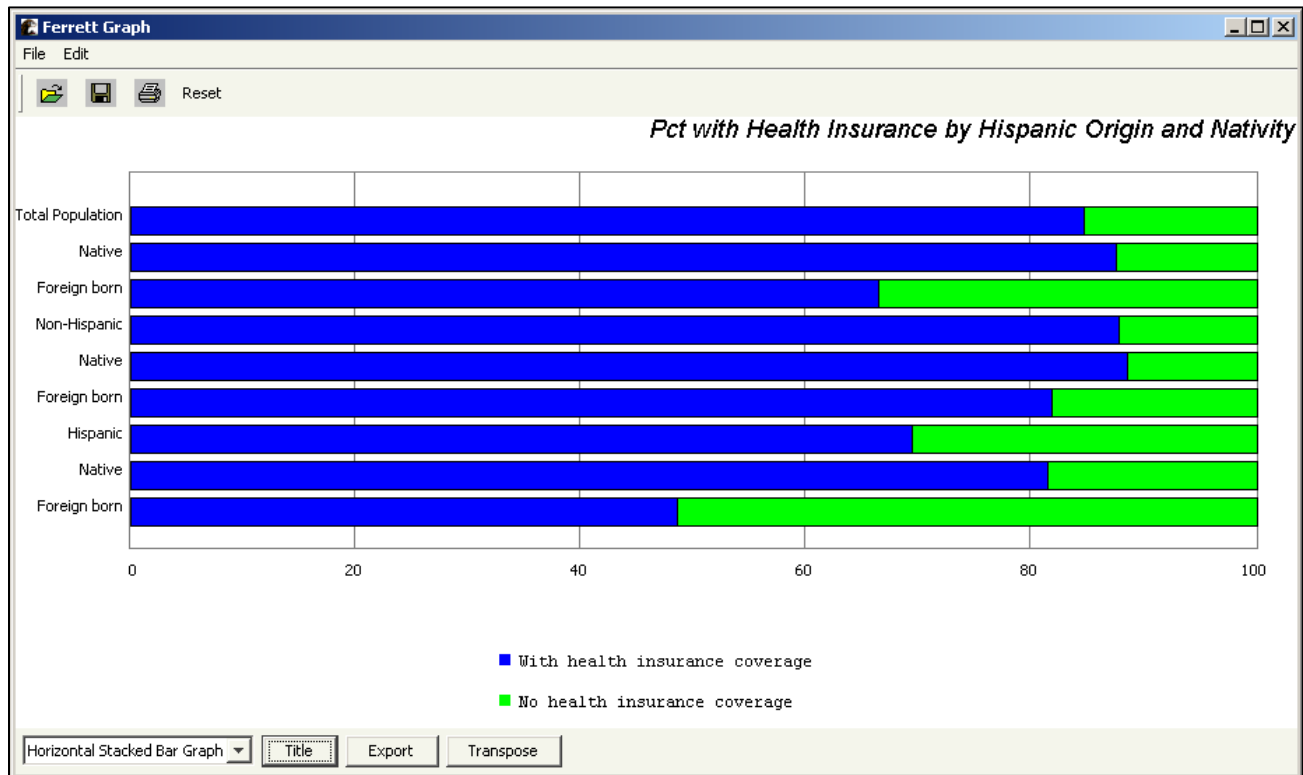


Figure 5-14: Graph Modified by Changing Type

Graph Types

Horizontal Bar Graph

The Horizontal Bar Graph is a comparative representation of data where values for each variable are shown as separate individual bars. This graph is arranged horizontally, as shown in Figure 5-13.

Horizontal Stacked Bar Graph

The Horizontal Stacked Bar Graph is similar to the horizontal bar graph in that values for variables are represented by individual bars. However, in the stacked bar graph, the values for each variable are shown as one continuous bar that adds to the total for the variable. This graph is also arranged horizontally, as shown in Figure 5-14.

Vertical Graph

Figure 5-15 shows the data in Figure 5-6 represented as a Vertical Bar Graph. Note that only some of the labels are printed due to space constraints on the horizontal axis. A work-around for this issue is to abbreviate the labels so that all data is identified.

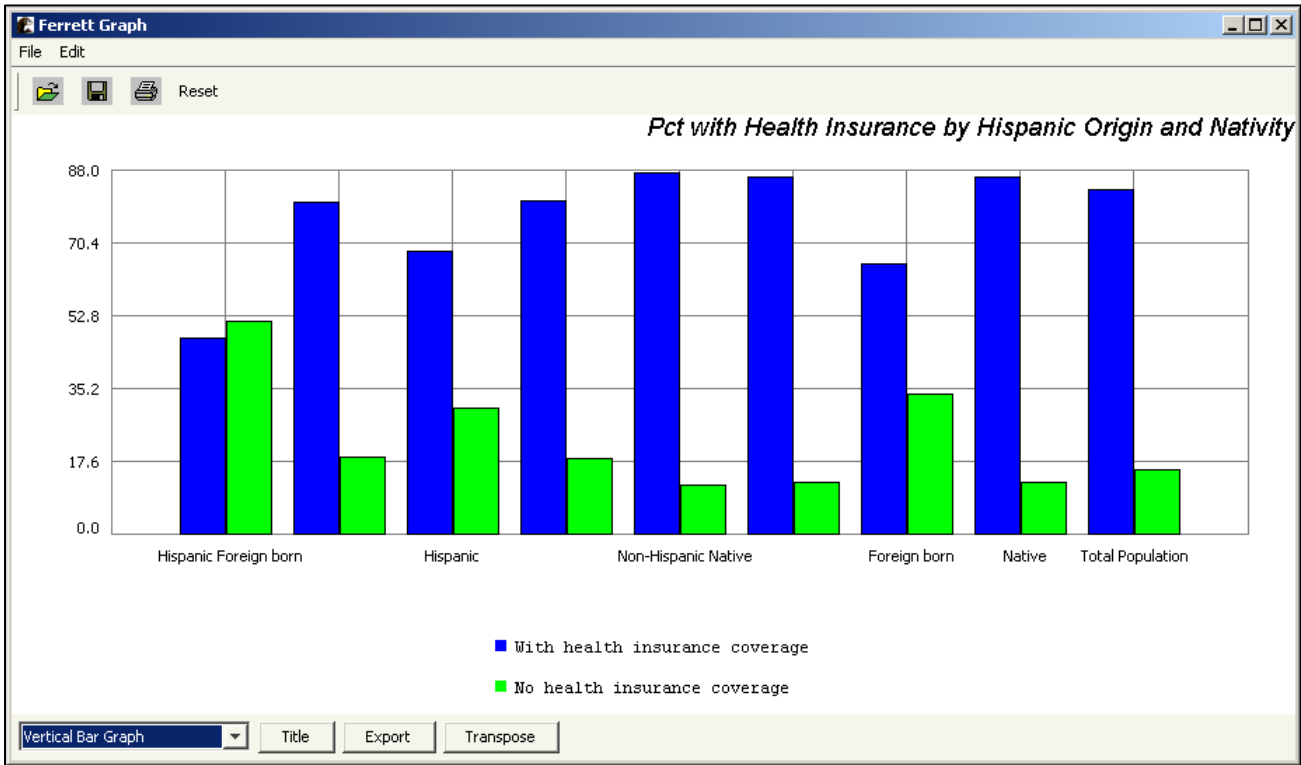


Figure 5-15: Vertical Bar Graph

Vertical Stacked Bar Graph

Figure 5-16 is identical in configuration to the Horizontal Stacked Bar Graph, arranged vertically.

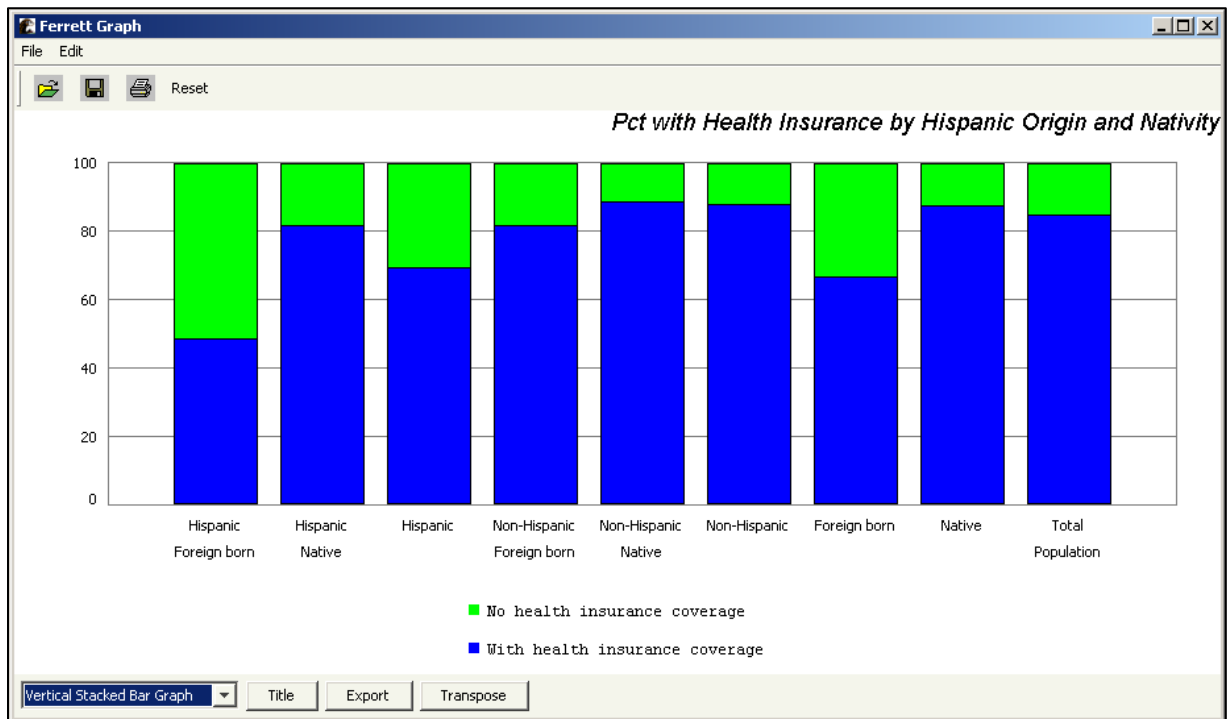


Figure 5-16: Vertical Stacked Bar Graph

Line Graph

In order to define and demonstrate the creation of a line graph, the change between 2001 and 2011 in participation in the labor force for the “older population,” defined as the 40-69 age range for this example, will be examined. Table 5-1 below shows the percent in the labor force between March 2001 and March 2011 for various age groups, taken from the CPS March Supplement.

	C1	C2	C3	C4	C5	C6	C7	C8	C9
R1		Total AGE1	40 to 44 years	45 to 49 years	50 to 54 years	55 to 59 years	60 to 61 years	62 to 64 years	65 to 69 years
R2	Total INSTANCES	100.0	24.3	24.2	21.4	16.3	4.8	4.8	4.1
R3	Mar 2011	100.0	21.2	22.4	21.6	17.8	5.7	6.3	5.1
R4	Mar 2010	100.0	21.3	23.2	21.9	17.4	5.5	5.9	4.7
R5	Mar 2009	100.0	22.0	23.5	21.8	17.2	5.6	5.3	4.5
R6	Mar 2008	100.0	22.7	23.9	21.6	17.1	5.5	4.9	4.4
R7	Mar 2007	100.0	23.6	24.2	21.5	17.0	4.9	4.7	4.1
R8	Mar 2006	100.0	24.4	24.7	21.2	16.8	4.4	4.8	3.7
R9	Mar 2005	100.0	25.4	24.8	21.1	16.1	4.5	4.4	3.7
R10	Mar 2004	100.0	25.9	24.9	21.0	15.8	4.5	4.1	3.9
R11	Mar 2003	100.0	26.5	24.9	21.1	15.4	4.3	4.1	3.7
R12	Mar 2002	100.0	27.3	25.2	21.3	14.7	3.9	4.2	3.5
R13	Mar 2001	100.0	28.7	25.5	21.0	13.7	3.8	3.8	3.5

Table 5-1: Percent Participation in Labor Force for Select Age Groups Based on CPS March Supplement 2001 - 2011

For this example, only the elderly age groups represented in the last three columns (ages 60-61, 62-64, and 65-69) will be examined over time using the DataFerrett Line Graph.

Table 5-2 below shows which cells must be highlighted to graph the three age groups over the time period. Notice that the numbers in row R2 are not included because they represent an average labor force participation rate across the entire period. Figure 5-17 below shows the resulting graph. This graph type is particularly useful for analyzing change over time, and it allows the user to examine the changes for several groups or characteristics simultaneously.

	C1	C2	C3	C4	C5	C6	C7	C8	C9
R1		Total AGE1	40 to 44 years	45 to 49 years	50 to 54 years	55 to 59 years	60 to 61 years	62 to 64 years	65 to 69 years
R2	Total INSTANCES	100.0	24.3	24.2	21.4	16.3	4.8	4.8	4.1
R3	Mar 2011	100.0	21.2	22.4	21.6	17.8	5.7	6.3	5.1
R4	Mar 2010	100.0	21.3	23.2	21.9	17.4	5.5	5.9	4.7
R5	Mar 2009	100.0	22.0	23.5	21.8	17.2	5.6	5.3	4.5
R6	Mar 2008	100.0	22.7	23.9	21.6	17.1	5.5	4.9	4.4
R7	Mar 2007	100.0	23.6	24.2	21.5	17.0	4.9	4.7	4.1
R8	Mar 2006	100.0	24.4	24.7	21.2	16.8	4.4	4.8	3.7
R9	Mar 2005	100.0	25.4	24.8	21.1	16.1	4.5	4.4	3.7
R10	Mar 2004	100.0	25.9	24.9	21.0	15.8	4.5	4.1	3.9
R11	Mar 2003	100.0	26.5	24.9	21.1	15.4	4.3	4.1	3.7
R12	Mar 2002	100.0	27.3	25.2	21.3	14.7	3.9	4.2	3.5
R13	Mar 2001	100.0	28.7	25.5	21.0	13.7	3.8	3.8	3.5

Table 5-2: Age Groups Selected for Examination

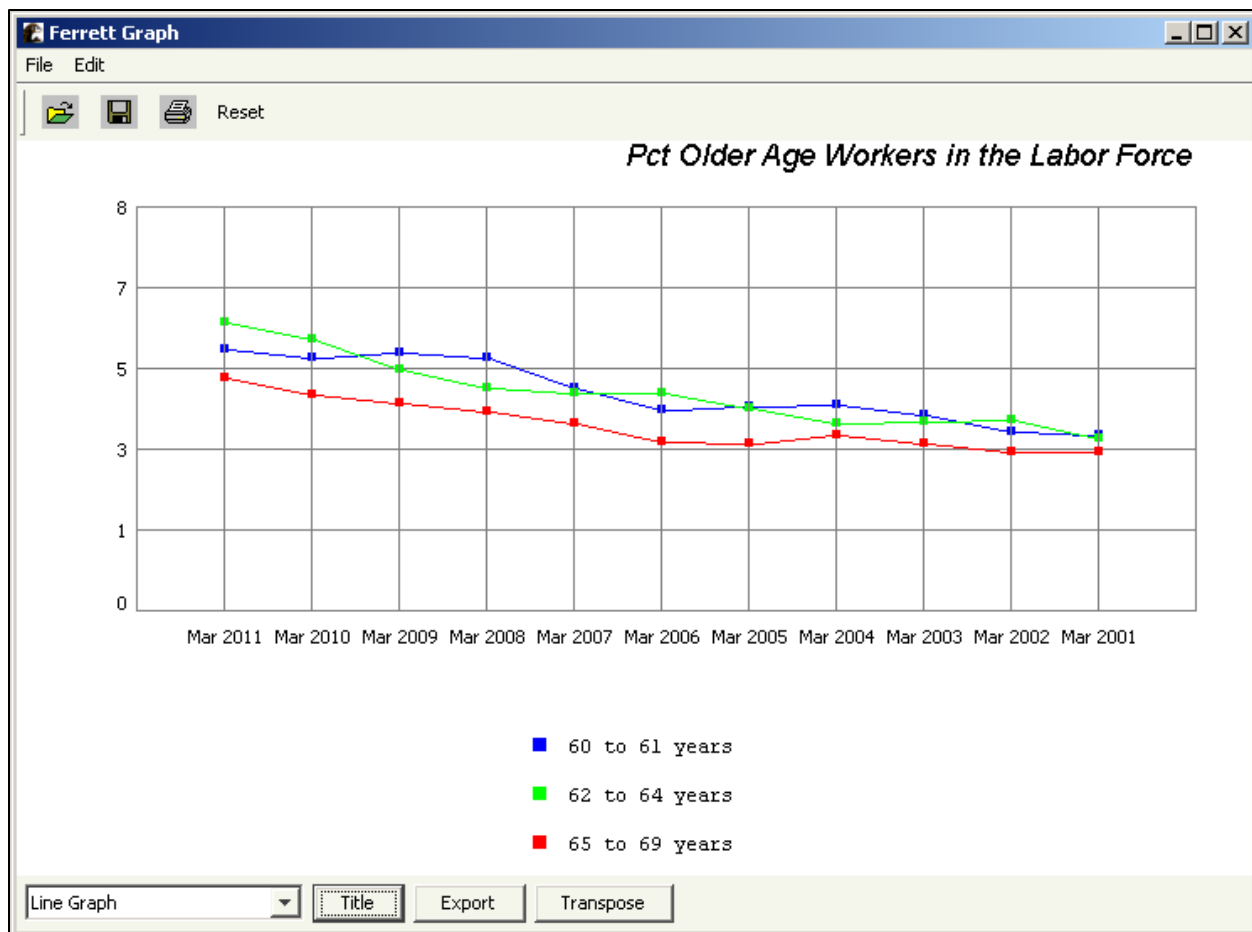


Figure 5-17: Line Graph Based on Selected Data

The graph that is most appropriate for this examination is the line graph. Figure 5-18 below shows this graph after you choose the line graph type, click on “reverse axis” in the edit menu of the graph window, and give the graph a title.

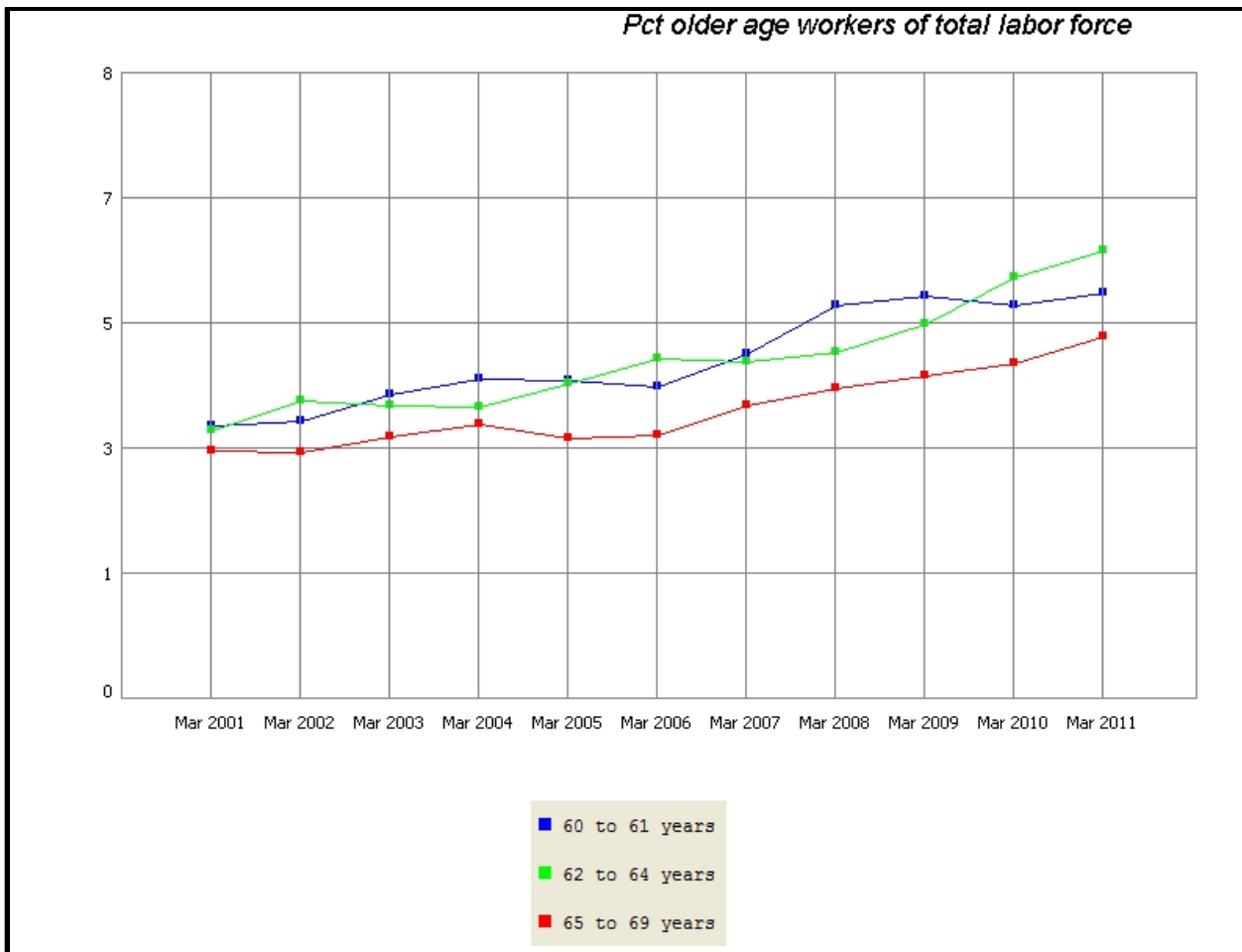


Figure 5-18: Modified Line Graph

Pyramid Graph

Referring to Table 5-1, notice that the two columns, 40-44 years old and 45-49 years old, show a reduction in the percent in the labor force from 2001 to 2011. Figure 5-19 uses the Pyramid Graph to show this reduction over time.

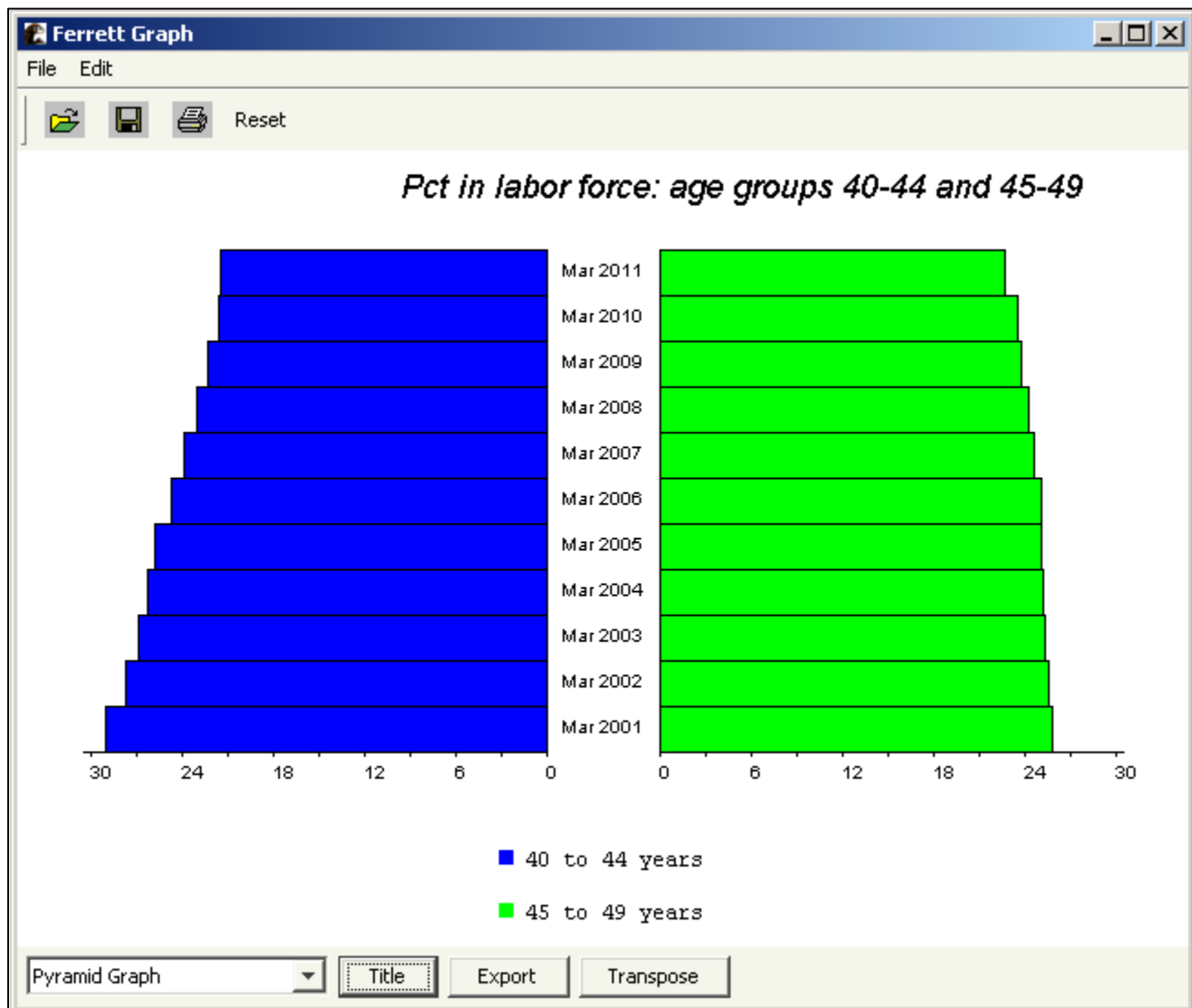


Figure 5-19: Pyramid Graph

NOTE: Caution is advised when using the Pyramid graph type for more than two groups, since the graph can be confusing to interpret.

Pyramid Percent Graph

The Pyramid Percent graph should only be used when the percent display option in the table is turned off. If this graph type is chosen when the percent option is on, the results might be misleading.

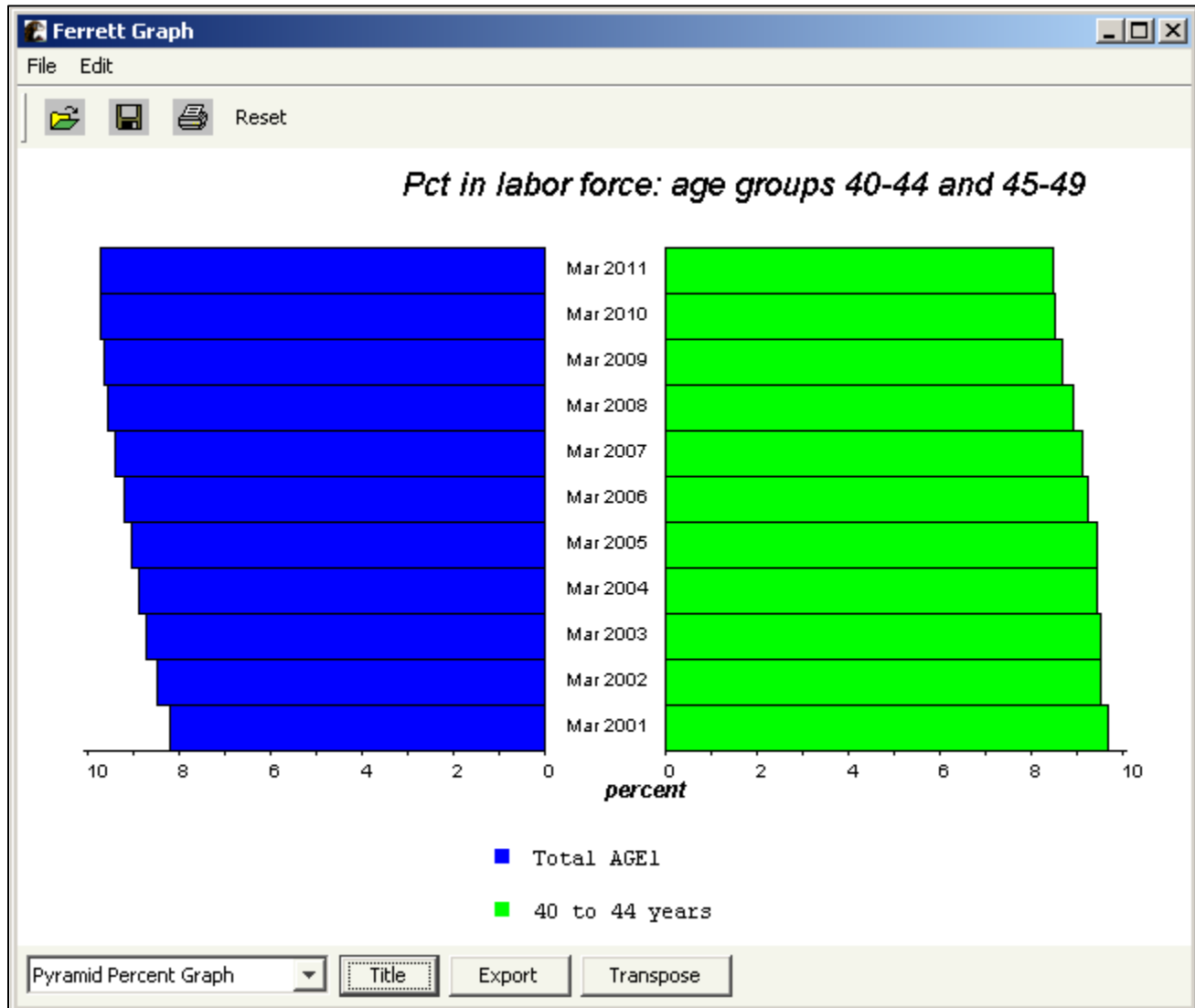


Figure 5-20: Pyramid Percent Graph

Pie Graph

This graph type, often called a “Pie Chart,” is widely used to represent the percentage shares of a total within a single time period. Using our current example from the eleven years of CPS March Supplement data (Table 5-1), a single Pie Graph for the distribution of percent in the labor force across all age groups for 2011 can be created and is shown below.

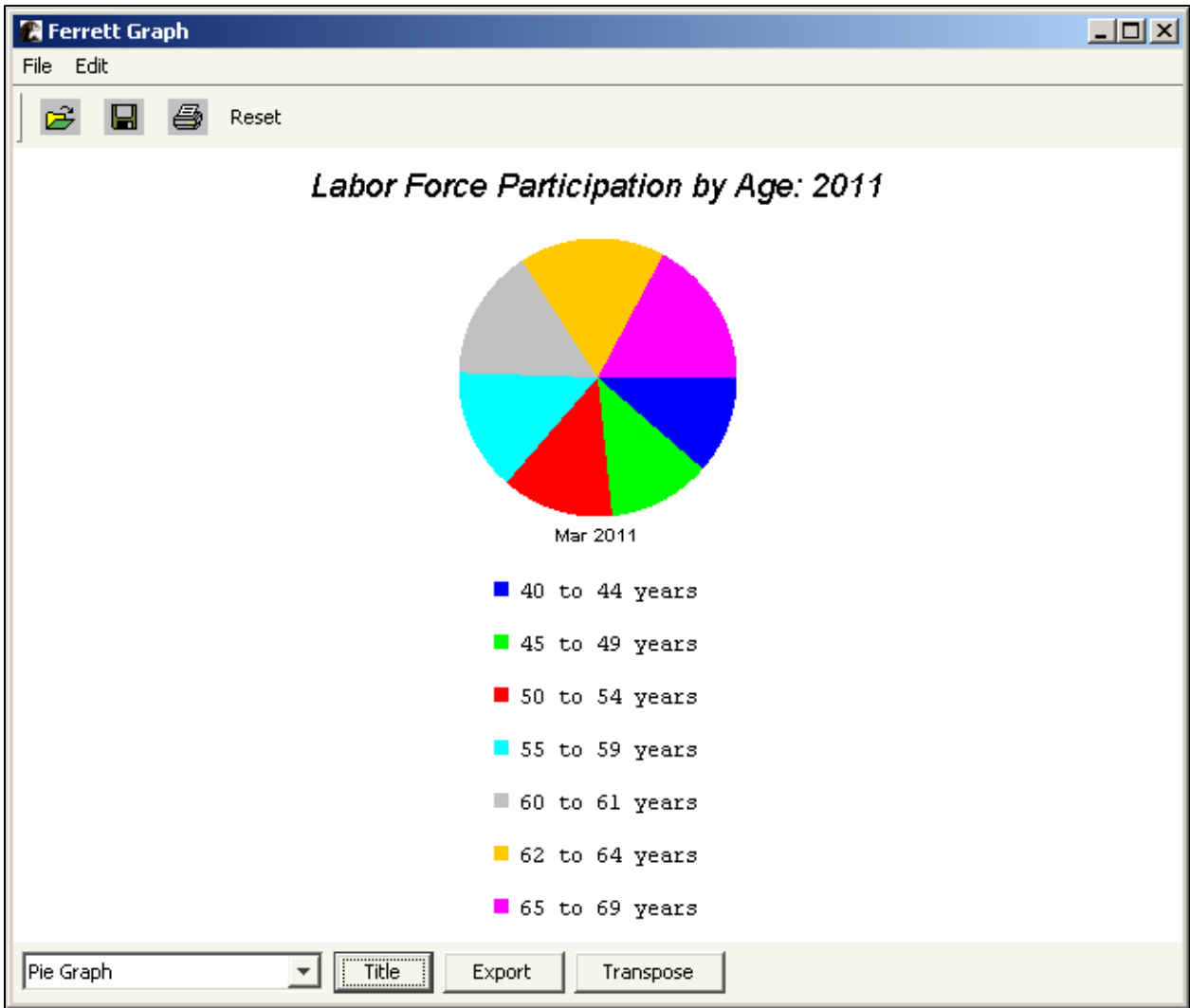


Figure 5-21: Single Pie Graph

The pie graph can also be used to compare the first year of the time interval with the last year, as shown below in Figure 5-22.

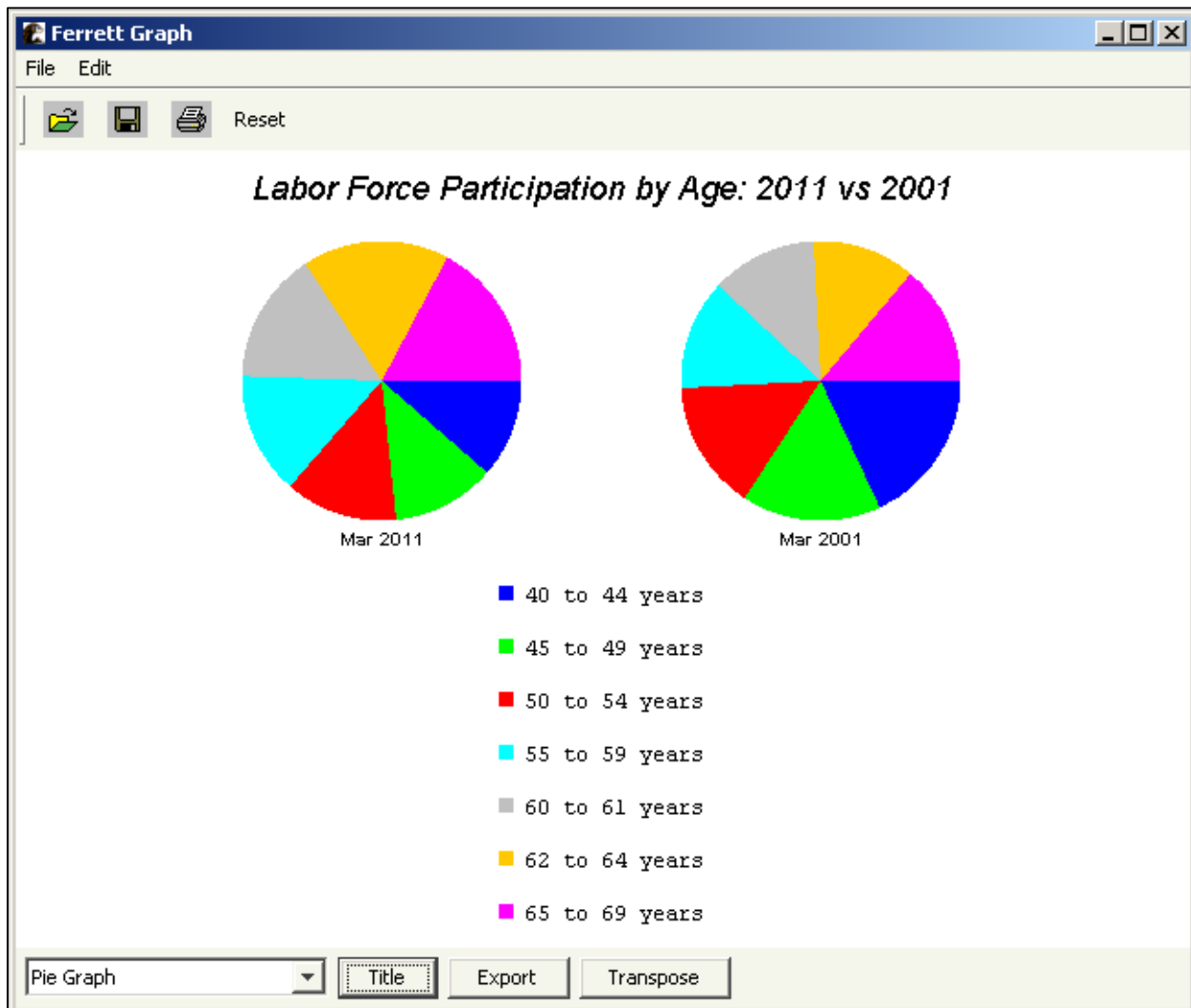


Figure 5-22: Comparative Pie Graphs

Area Graph

An Area Graph allows you to see the change in size over time of several groups on the same graph. In Figure 5-23 below, the area graph shows that the 40-44 age group percent shrinks over time more than the 45-49 age group. The group that is largest at the end of the time period remains in front, and the other group is shown behind the first group.

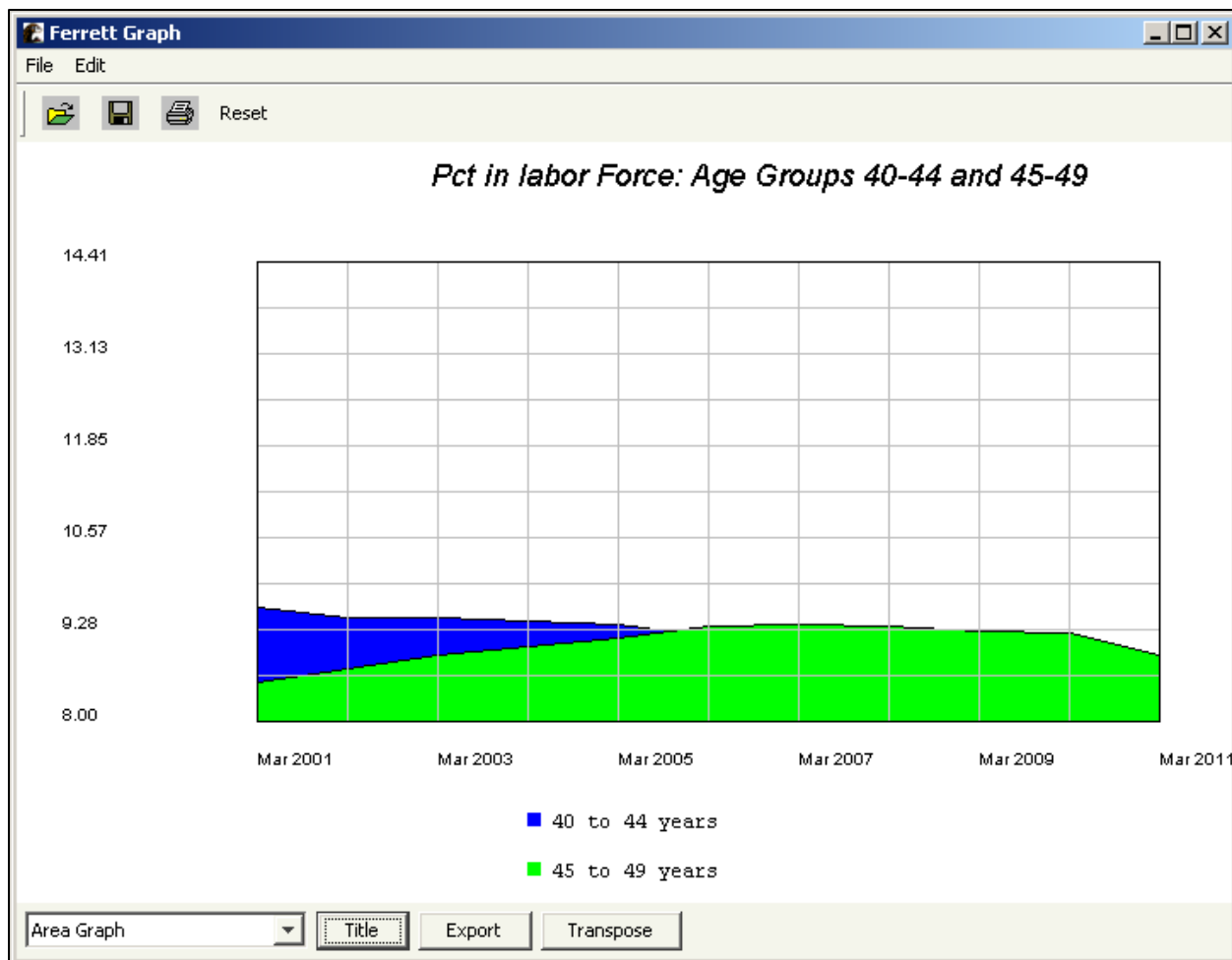


Figure 5-23: Area Graph

Dynamic Area Graph

This graph type is a variation of the area graph. Instead of one group in front of another, the groups are stacked on top of each other. Note that the years are shown on top at the top for this graph type.

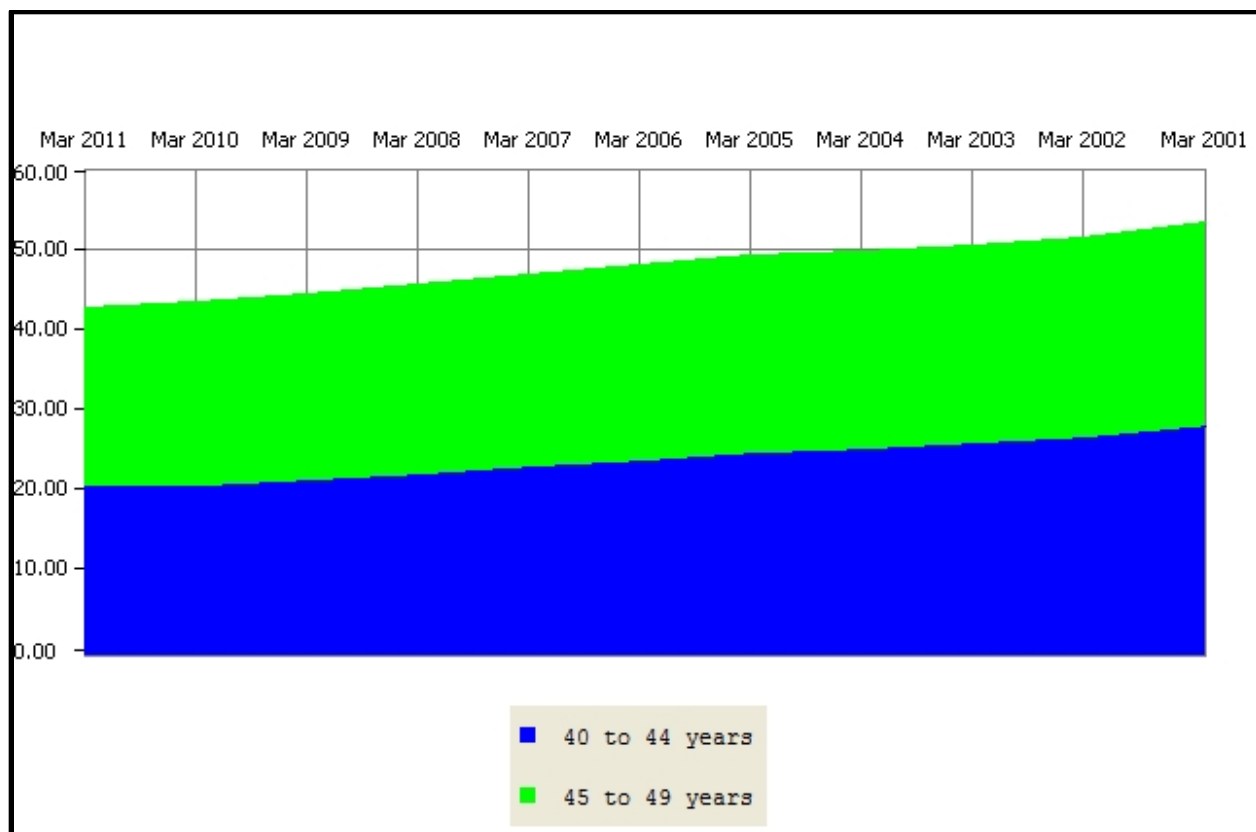


Figure 5-24: Dynamic Area Graph

CHAPTER 6: MAKE A MAP

“Chapter 4: Make a Table” briefly discussed the map icon located on the tool bar of the “Ferrett Tabulation Window” (Figure 6-1). This chapter explains the mapping function and demonstrates creating data maps using data from the 2006-2010 5-Year Summary File in the American Community Survey.



Figure 6-1: Map Creation Button

DataFerrett offers various options for customizing maps that present data in the most effective manner. These options include:

- Change Map Size
- Change Legend Class
- Change Color Scheme
- Change Range Values
- Specify Range Intervals

Example 6-1: Map Comparison of Citizenship Status

A description of each option follows the example below.

The following example demonstrates how to create a map that compares citizenship status throughout the United States by using the 2006-2010 Summarized Data from the American Community Survey.

NOTE: The map button becomes active once data is added to and highlighted in the spreadsheet that is created in “Step 2: DataBasket/Download/Make a Table.”

Example 6-1: Map Comparison of Citizenship Status		
Steps	What to do	Why do it?
1.1 1.2 1.3	<p>Select variables related to U.S. citizenship status (B05001) from the ACS Summarized Data – 5 Year Summary File for 2006-2010. Highlight all variables, click “Browse/Select Highlighted Variables” and add the variables to the DataBasket.</p> <p>In the “Required Variables” window that opens next, double click “Selectable Geographies” to open the Ferrett Geography Codebook. Highlight “State” under Type of Geography and “State (FIPS Code)” under Hierarchies; click “Use Hierarchy.” Drag “Select All” to the right panel and click “Finish” to add the selected geography to the DataBasket.</p> <p>The variables that are now in the DataBasket are shown in Figure 6-2.</p> <p>Select the “Step 2” tab and click “Make a Table.”</p>	Making a table is the first step in the creation of a map.
2	<p>Place variables in the spreadsheet and click “Go Get Data.” Highlight all variables showing U.S. Citizenship and drop into cell C2. Highlight the column furthest to the right and name it “US Citizen.” In the formula bar (column still highlighted) type “=SUMC3:C6.” To create the layout shown in Figure 6-4, hide columns C3 through C6 by highlighting them and selecting “Hide” from the “Edit” menu. Highlight column C9 and type in “Pct Citizen,” go to the formula bar and type in “=COMP(C8/C2*100).”</p>	Refer to <i>Chapter 4: Make a Table, Example 4-1, Steps 2 and 3</i> . Your goal is to create a map based on two categories: “Not a U.S. Citizen” and “Total U.S. Citizens.” Since there are four categories encompassing citizenship, you need to simplify it by creating two categories. You also want to change the final calculation to percentages to simplify the creation of the map.
3	<p>Highlight the data in column C10 to obtain percentages of citizens for each State. Afterwards, select cell C10, R1 for custom label.</p>	You cannot get the percentages to create the map unless you limit your selection to the data cells (selecting the column will not give you the results). Furthermore, you must select the custom label cell in order

		for the title to appear at the top of the map.
--	--	--

Step 1

The screenshot shows the DataBasket web application interface. At the top, there are three tabs: "Introduction" (with a raccoon icon), "Step1 : Select Dataset & Variable" (active), and "Step2 : DataBasket/Download/Make A Table". Below the tabs, there is a text box that says "Review your variables then go back to select more variables or go on to get data". To the right of this text are three icons: a "Download" icon (a raccoon with a stack of papers), a "Make A Table" icon (a raccoon with a stack of papers), and an information icon (a lowercase 'i'). Below these icons is a table titled "Current Query Variables from ACS Summarized Data (5-Year Summary File):".

Name	Variable Label	Availability
B05001_001E	Total:	2005-2009 - current
B05001_002E	U.S. citizen, born in the United States	2005-2009 - current
B05001_003E	U.S. citizen, born in Puerto Rico or U.S. Island Areas	2005-2009 - current
B05001_004E	U.S. citizen, born abroad of American parent(s)	2005-2009 - current
B05001_005E	U.S. citizen by naturalization	2005-2009 - current
B05001_006E	Not a U.S. citizen	2005-2009 - current
COMPONENT	Geographic Component	2005-2009 - current
GEOG-101	FIPS State Code	2005-2009 - current

Figure 6-2: DataBasket

Step 2

Ferrett Tabulation

File Edit Format View Options Help

GO Get Data

Col C4: AGG(B05001_003E)=0

Pivot(s) can be dropped on pivot image above R1.

	C2	C3	C4	C5	C6	C7	C8
R1	Total:	U.S. citizen, born in the United States	U.S. citizen, born in Puerto Rico or U.S. Island Areas	U.S. citizen, born abroad of American parent(s)	U.S. citizen by naturalization	Not a U.S. citizen	
R2	?		?	?	?	?	?
R3	?		?		?	?	?
R4	?		?	?	?	?	?
R5	?		?	?	?	?	?
R6	?		?	?	?	?	?
R7	?		?	?	?	?	?
R8	?		?	?	?	?	?
R9	?		?	?	?	?	?
R10	?		?	?	?	?	?
R11	?		?	?	?	?	?
R12	?		?	?	?	?	?
R13	?		?	?	?	?	?
R14	?		?	?	?	?	?
R15	?		?	?	?	?	?
R16	?		?		?	?	?
R17	?		?	?	?	?	?
R18	?		?	?	?	?	?
R19	?		?	?	?	?	?
R20	?		?	?	?	?	?
R21	?		?	?	?	?	?
R22	?		?	?	?	?	?
R23	?		?	?	?	?	?
R24	?		?	?	?	?	?
R25	?		?	?	?	?	?
R26	?		?	?	?	?	?
R27	?		?	?	?	?	?
R28	?		?	?	?	?	?
R29	?		?	?	?	?	?
R30	?		?	?	?	?	?
R31	?		?	?	?	?	?
R32	?		?	?	?	?	?

Figure 6-3: Drag and Drop Variables into Spreadsheet

Step 2, cont.

Ferrett Tabulation

File Edit Format View Options Help

GO Get Data

Formula Bar

Pivot(s) can be dropped on pivot image above R1.

	C1	C2	C3	C4	C5
R1		Total:	U.S. citizen, bom in the United States	U.S. citizen, bom in Puerto Rico or U.S. Island Areas	U.S. citizen, bor
R2	Alabama	4,712,651		4,519,341	5,801
R3	Alaska	691,189		626,428	3,779
R4	Arizona	6,246,816		5,290,576	11,629
R5	Arkansas	2,872,684		2,731,915	2,893
R6	California	36,637,290		26,256,871	80,605
R7	Colorado	4,887,061		4,341,972	7,934
R8	Connecticut	3,545,837		2,956,434	94,153
R9	Delaware	881,278		794,916	7,460
R10	District of Columbia	584,400		500,014	1,814
R11	Florida	18,511,620		14,392,415	382,139
R12	Georgia	9,468,815		8,447,606	31,498
R13	Hawaii	1,333,591		1,058,649	14,173
R14	Idaho	1,526,797		1,423,272	1,622
R15	Illinois	12,745,359		10,885,810	53,952
R16	Indiana	6,417,398		6,095,169	8,803
R17	Iowa	3,016,267		2,877,468	2,163
R18	Kansas	2,809,329		2,609,061	3,914
R19	Kentucky	4,285,828		4,127,506	4,115
R20	Louisiana	4,429,940		4,246,402	4,972
R21	Maine	1,327,665		1,270,350	1,607
R22	Maryland	5,696,423		4,872,701	16,815
R23	Massachusetts	6,477,096		5,378,552	106,977
R24	Michigan	9,952,687		9,300,106	12,242
R25	Minnesota	5,241,914		4,844,621	4,116
R26	Mississippi	2,941,991		2,863,856	2,487
R27	Missouri	5,922,314		5,669,226	4,814

Figure 6-4: Spreadsheet Populated with Data

Step 2, cont.

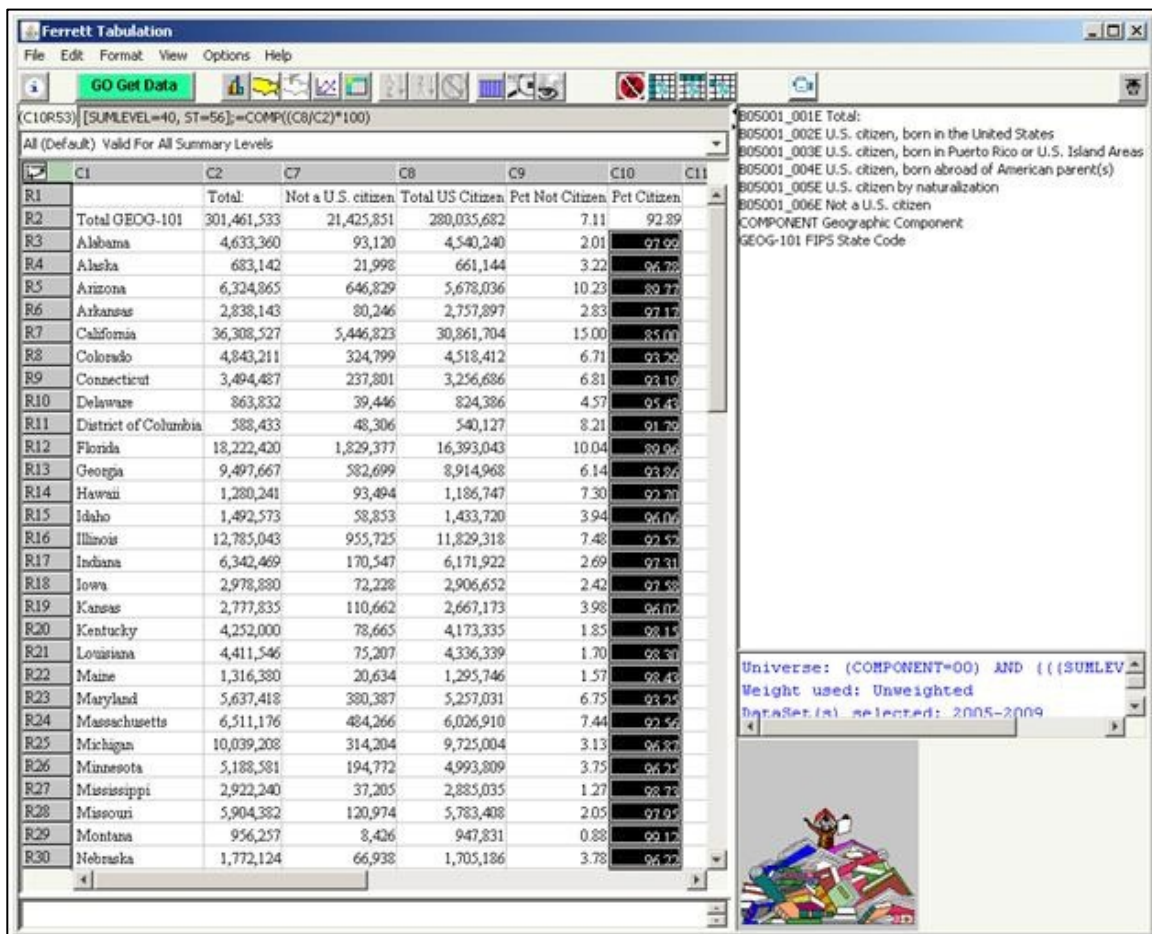


Figure 6-4: Select Data Column

Step 3

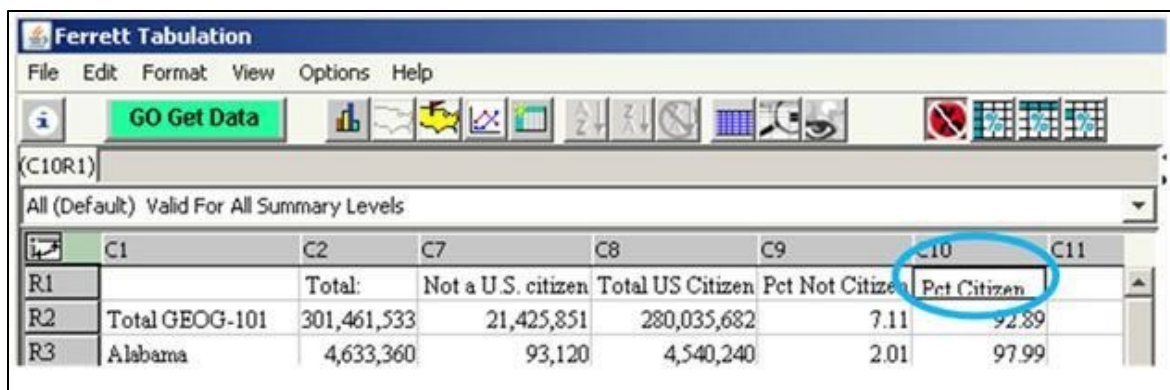


Figure 6-5: Select Cell for Custom Label

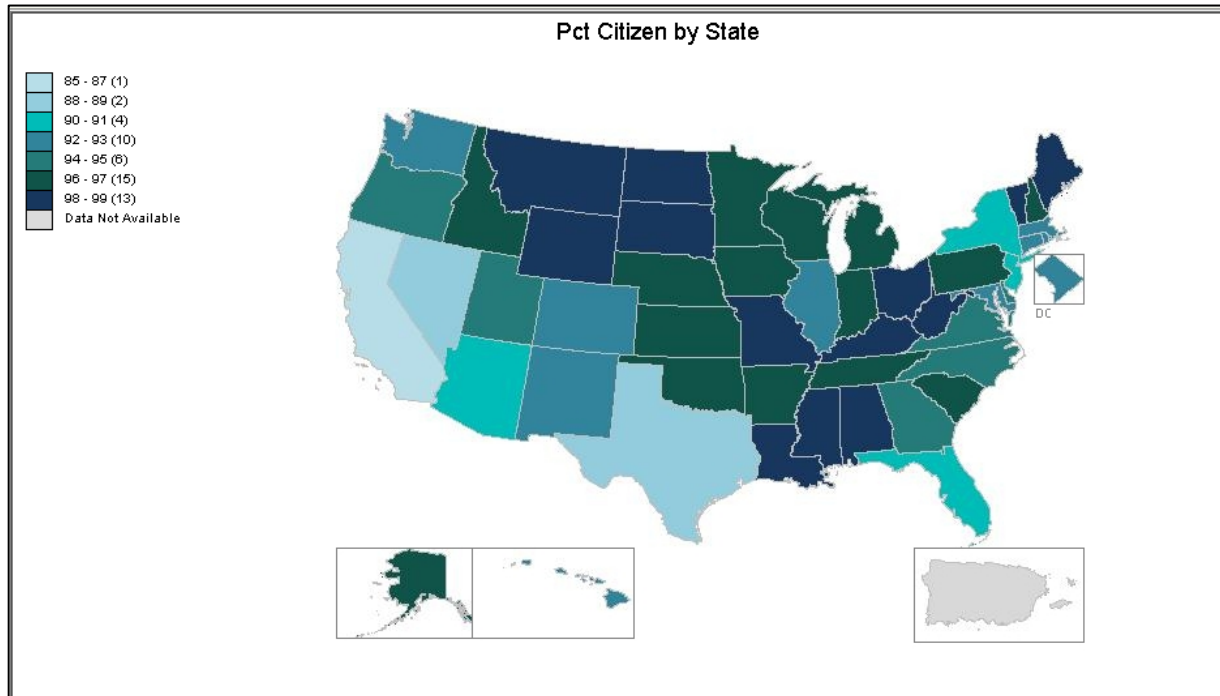


Figure 6-6: Map Generated

Customizing a Map

Changing Map Window Size

To change the size of the map window, click on View → Map Size. Type in the desired width and height of the map window, then click 'OK'. This feature is useful when saving the map image as a .jpg file or printing the screen for use in another document such as PowerPoint or Microsoft Word.

Changing Legend Classes

To change the Legend Classes (or Ranges) from the default value of seven, click on View → Legends → Legend Classes. Next, type in the desired number of legend classes, or ranges, and click 'OK'. The map image will automatically redraw based on the new number of legend classes.

Changing Color Schemes

To change the color schemes of the map, click on View → Legend Types. The image shown in Figure 6-7 below will appear, providing the map types and color schemes that are available.

Once you have selected the desired color scheme, click 'OK'. The map image will automatically redraw based on the new legend type or color scheme.



Figure 6-7: Legend Types

Changing Minimum and Maximum Values for the Ranges

To change the displayed minimum and maximum values on the map, click on View → Value Range. Afterwards, type in the desired Start Value and End Value and click 'OK'. The map will automatically redraw based on the new minimum and maximum (start and end) values.

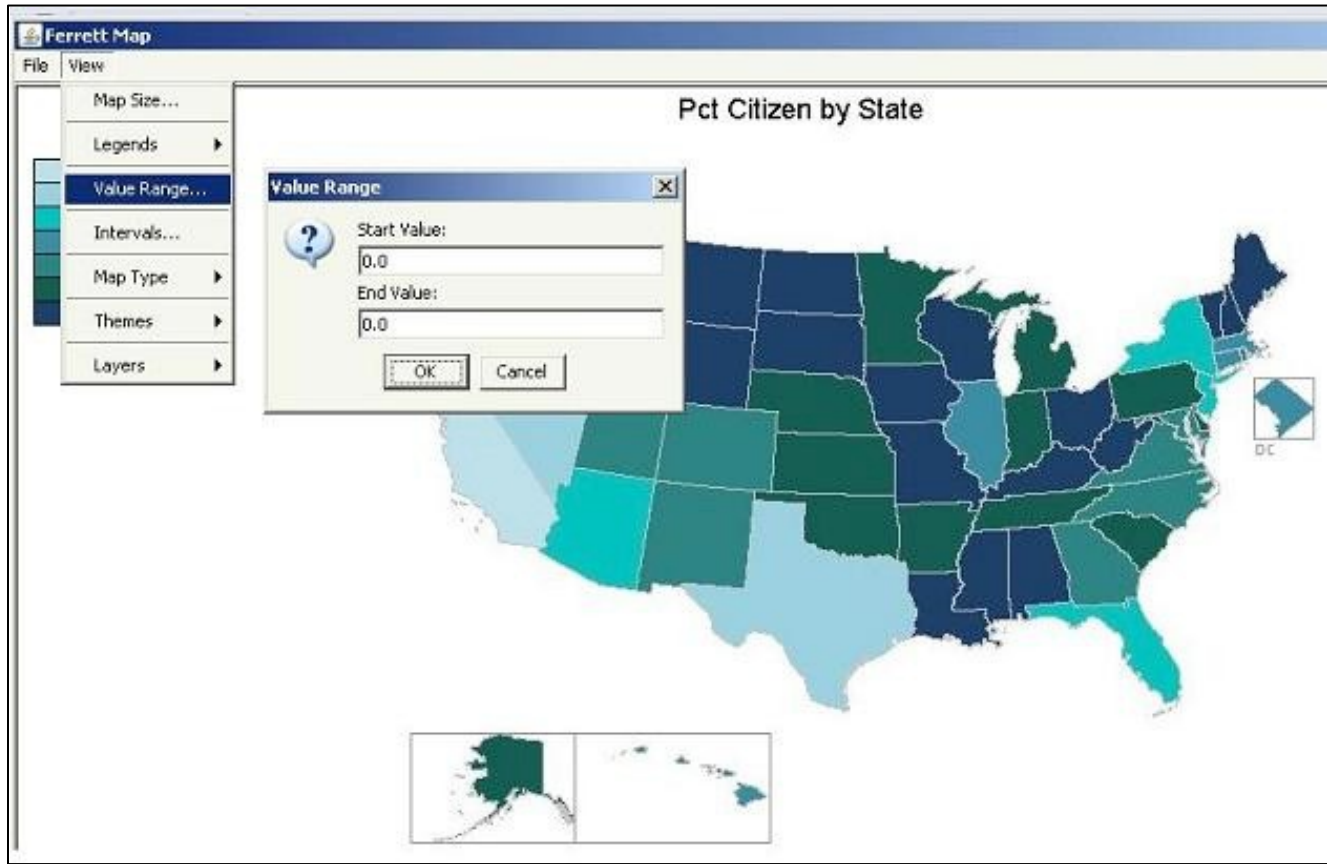


Figure 6-8: Selecting Value Ranges

Customizing Interval Breaks in the Legend

To customize the interval break-outs for each range, click on View → Intervals. A window will pop up that allows you to type in specified ranges (Figure 6-9).

Interval:	From:	To:
1	8500	8706
2	8706	8912
3	8912	9118
4	9118	9324
5	9324	9530
6	9530	9736
7	9736	9936

Figure 6-9: Specify intervals

Changing Map Display and Legend to Quantile Intervals

To change the map display and legend from the default Equal Interval to Quantile, click on View → Map Type, Select Quantile. The map will automatically redraw to reflect the new quantile thematic ranges:



Figure 6-10: Changing Map Display

Adding Layer Boundaries

Depending on the level of geography being mapped; i.e., states, counties, census tracts, etc. that are available for selected datasets in DataFerrett, there is an option to turn on or off specific boundary layers (Figure 6-11).

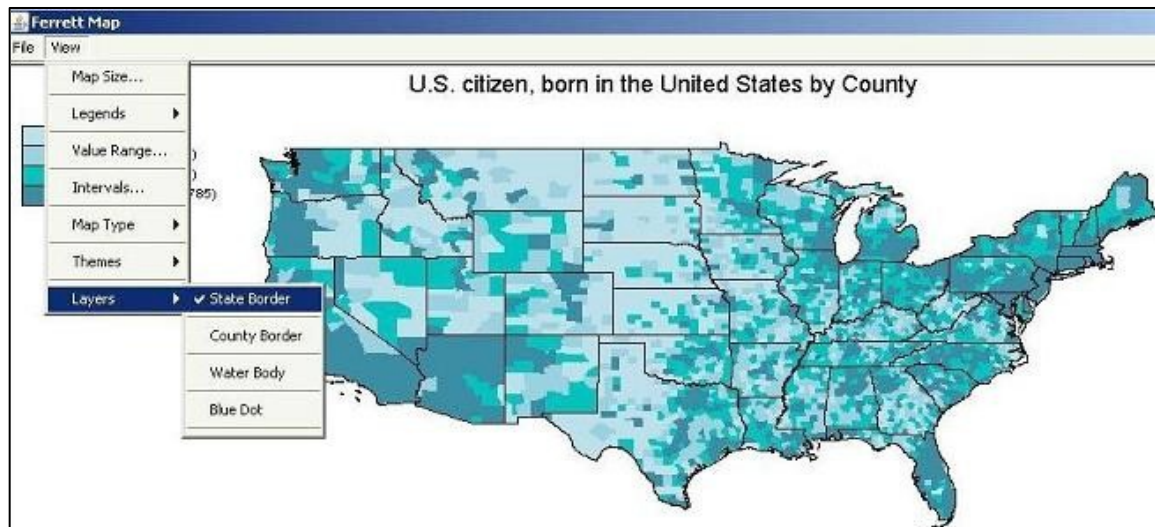


Figure 6-11: Add Boundaries

Along with basic geographic boundary layers like states, counties, census tracts, etc. that can be turned on or off, there is also the opportunity to turn on a point layer (“Blue Dot”) where appropriate. The “Blue Dot” layer displays address points assuming that a dataset containing appropriate latitude and longitude information has been tabulated and is being added to an existing thematic map.

CHAPTER 7: DOWNLOAD AN EXTRACT

In addition to providing options for displaying and presenting data, the “Step 2: DataBasket/Download/Make a Table” tab in DataFerrett enables the user to extract variables in the DataBasket to an external file by use of the “Download” feature, highlighted in Figure 7-1 below. Because DataFerrett provides so many ways to define subsets of the full dataset, this feature is particularly useful when a small part of a very large dataset is needed. For example, by restricting the values in an age variable to a certain range, the user can download all variables of interest for the selected age range alone without considering an entire dataset.

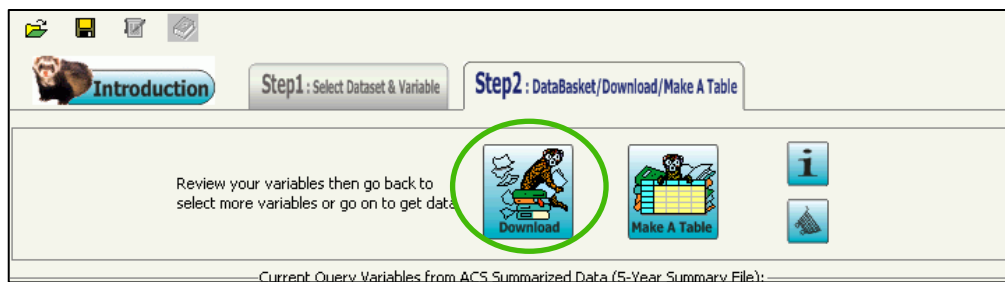


Figure 7-1: Download Icon in “Step 2” Screen

Example 7-1: Downloading Extracts from Select Dataset

Example 7-1: Downloading Extracts from Select Dataset		
Steps	What to do	Why do it?
1	We will be referring to the first example from Chapter 1 in which we added two variables from the 2009 ACS PUMS to our DataBasket.	The Decade of Entry and Hispanic Recode (limited to South America) contain variables applicable for this method.
2	We will also be adding all the state codes and all PUMA codes to the DataBasket.	
3	Click on the Step 2 tab and click on the Download icon. There are now four variables in your DataBasket (Figure 7-2).	After clicking on the Download tab, a window will appear containing choices for the layout and format of the file being downloaded. The window contains default settings for four download specification questions, which will be discussed individually.

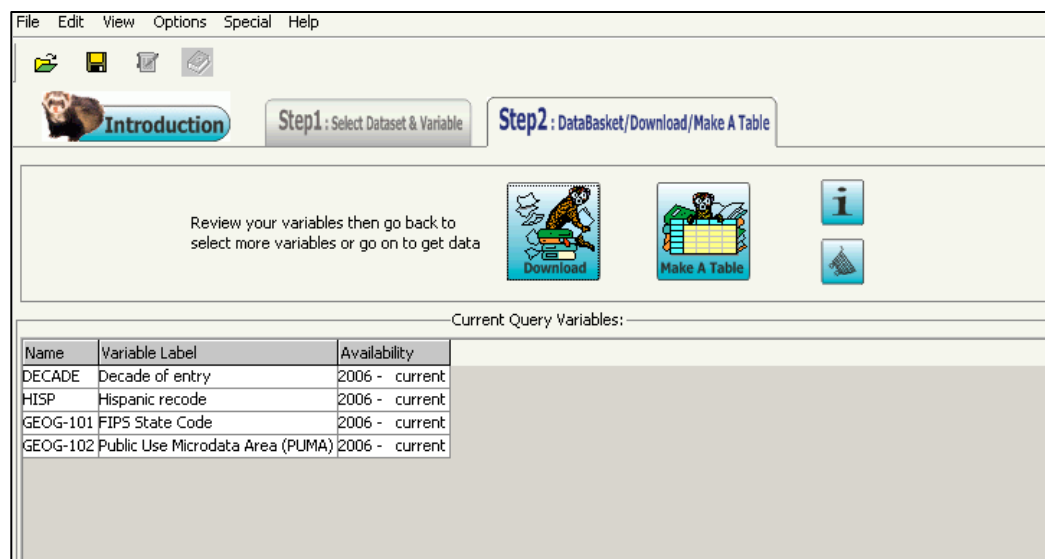


Figure 7-2: DataBasket After Adding States and PUMAs

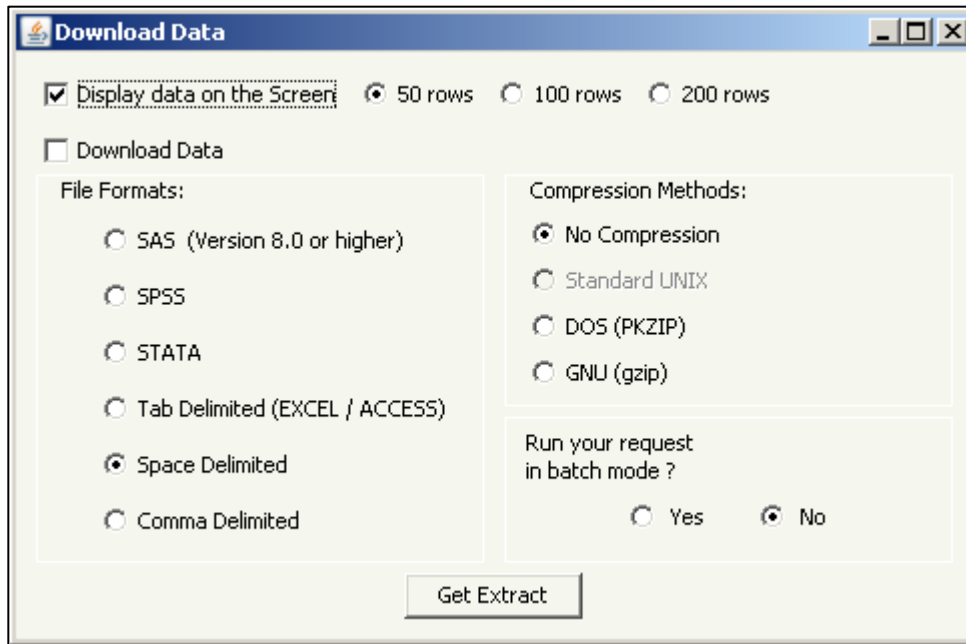


Figure 7-3: Download Data Specification Pop-up Window

Specification Question 1: Download or Display Only?

In order to limit the number of rows displayed to either 50, 100 or 200, click the radio button for the desired number of records to display. Do not check the box labeled “Download Data.” When you click on “Get Extract,” you will be sent to a web page named “DataFerrett Extraction Page”. (**NOTE:** This web page may be minimized on the method bar at the bottom of your screen.)

In the middle of this web page, you will see a file name similar to “Display File for 2009 douglas_hillmergq41e9611.txt.” By clicking on the file name (which is stored as a hyperlink), you will see the display of the first records in the extract file. Figure 7-4 below is an example of what you might see if you choose to restrict the number of records to display.

In order to display the complete file, check the box in front of “Download Data” on the DataFerrett Extraction Page. If this box is checked, it makes no difference which radio button is active for “Display data on the Screen,” as the entire file will be created and made available for download.

Obs	SERIALNO	DECADE	HISP	ST	PUMA
1	16	6	2	6	4300
2	16	5	2	6	4300
3	16	0	2	6	4300
4	16	0	2	6	4300
5	16	7	2	6	4300
6	22	0	1	6	1103
7	22	0	1	6	1103
8	62	6	1	6	2104
9	62	7	1	6	2104
10	62	0	1	6	2104
11	126	3	1	6	7601
12	126	0	1	6	7601
13	159	0	1	6	3302
14	159	0	1	6	3302
15	159	0	1	6	3302
16	159	0	1	6	3302
17	208	5	1	6	8004
18	208	5	1	6	8004
19	208	0	1	6	8004

Figure 7-4: Example of Display of First Records in an Extract File

Specification Question 2: What Data Format do You Want for the Downloaded Data?

All data extracts are downloaded as ascii text files. DataFerrett offers several options for download file format and layout. These options include “delimited” files with either a TAB character or a comma as the delimiter, or files with the fields in fixed starting positions on the record and a fixed length for each field. There are also format options that target three popular statistical software packages: SAS, SPSS and STATA.

For each of these options, an ascii data file and a file of program commands (in the syntax of the respective statistical software package) are extracted for download. Choosing your download format will be dependent upon your preferences for after-download file manipulation (in order to view all available file formats, check the box in front of “Download Data” before you click “Get Extract.”)

Selecting File Formats

The six choices under “File Formats” are listed below with a brief description of what the output will look like for each choice:

1. SAS (Version 8.0 or higher) – Text file with variables in fixed positions on each record will be output along with a file containing SAS code to read the file into a SAS dataset. **NOTE:** You will need to modify this code to work in your environment.
2. SPSS– Text file with variables in fixed positions on each record will be output along with a file containing SPSS code to read the file into a SPSS dataset. **NOTE:** You will need to modify this code to work in your environment.

3. STATA– Text file with variables in fixed positions on each record will be output along with a file containing STATA code to read the file into a STATA dataset. **NOTE:** You will need to modify this code to work in your environment.
4. Tab Delimited (EXCEL/ACCESS) – Text file with variables separated by a TAB character as the delimiter.
5. Space Delimited – Text file with variables in fixed positions on each record will be output along with a file containing record layout information for each record on the data file.
6. Comma Delimited - Text file with variables separated by comma as the delimiter.

Getting Extracts

After making a file format selection, select the Get Extract button to have the data downloaded to your computer. Below are images of the files that return with your data; some of these files will return with an additional statistical package command file.

SAS (version 8.0 or higher) Format: This format includes SAS input statements. As shown in Fig. 7-5 below, two files are returned. The file ending in “.asc” contains the data; the file ending in “.sas” contains the SAS commands needed to read the data file and output a SAS dataset. **NOTE:** The file of SAS commands must be edited in order to modify the “INFILE” statement to point to the actual name and location of the downloaded data file. An example of the record in the SAS command file with the “INFILE” statement is: infile ‘douglas_ hillmergq2fs0uz1.asc’ lrecl=64.

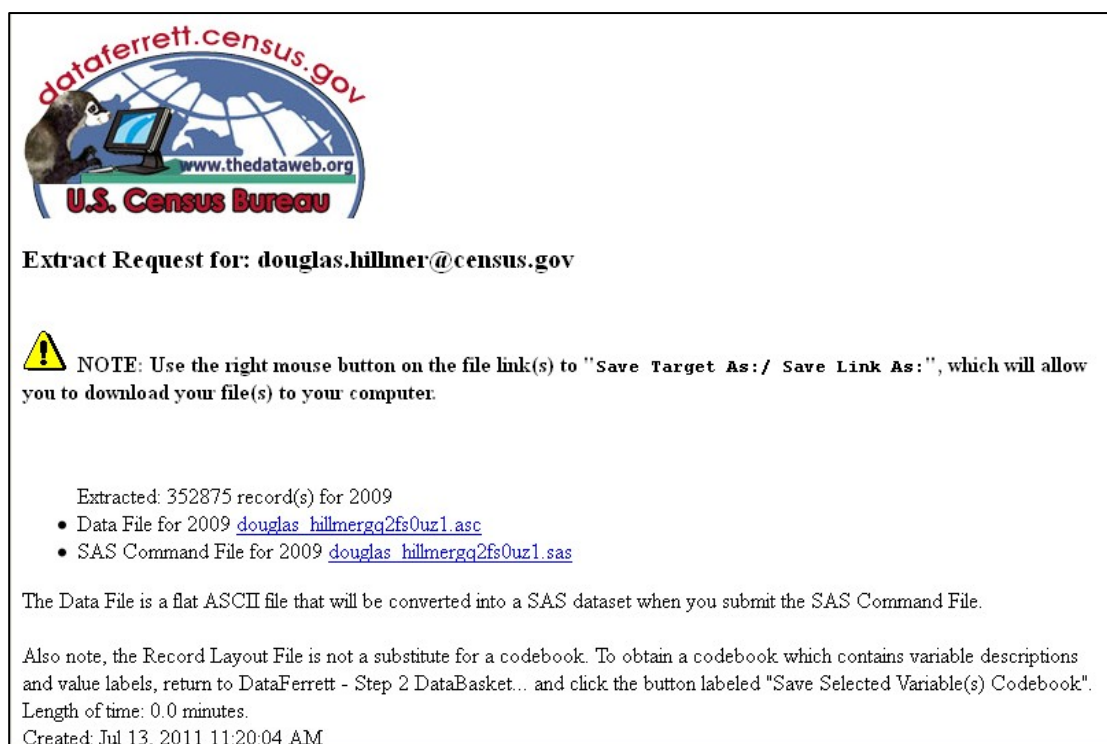


Figure 7-5: Extract Request Window for SAS Format

SPSS Format: Same as the SAS format with the file of commands in SPSS syntax.

STATA Format: Same as the SAS format with the file of commands in STATA syntax.

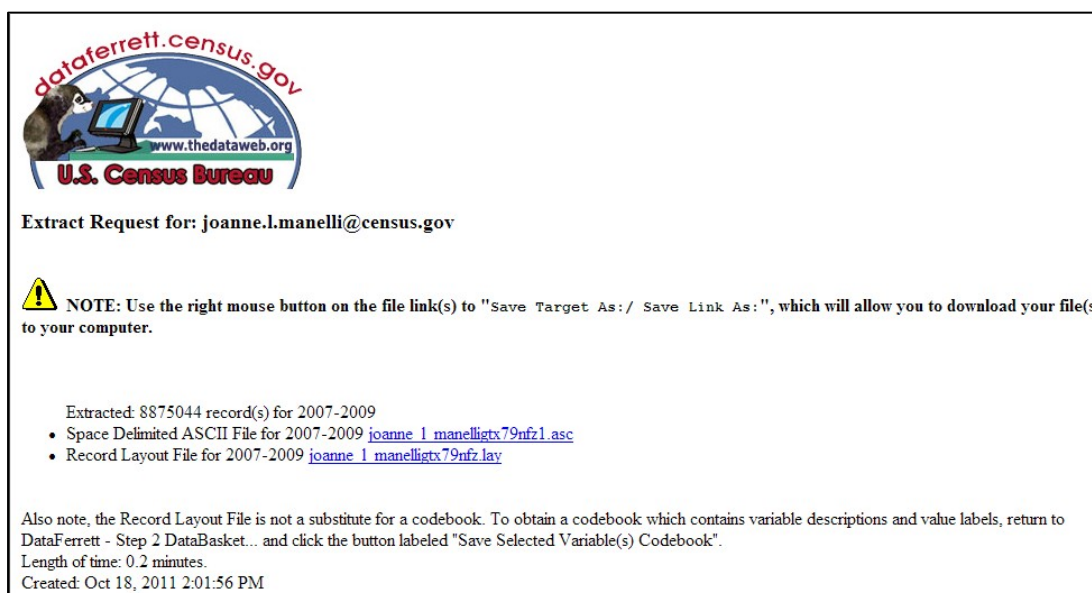
Tab Delimited (EXCEL / ACCESS): The example below displays the information returned on the extract results page for this format type:



The screenshot shows the DataFerrett interface for a Tab Delimited extract request. At the top is the DataFerrett logo with the URL dataferrett.census.gov and the U.S. Census Bureau logo. Below the logo, it says "Extract Request for: joanne.l.manelli@census.gov". A yellow warning icon is followed by a note: "NOTE: Use the right mouse button on the file link(s) to 'Save Target As:/ Save Link As:', which will allow you to download your file(s) to your computer." Below this is a bullet point: "Display File for 2009 [joanne l manelligtx3lojq1.txt](#)". At the bottom, it says: "Also note, the Record Layout File is not a substitute for a codebook. To obtain a codebook which contains variable descriptions and value labels, return to DataFerrett - Step 2 DataBasket... and click the button labeled 'Save Selected Variable(s) Codebook'." The length of time is 0.0 minutes and it was created on Oct 18, 2011 12:19:07 PM.

Figure 7-6: Extract Request for Tab Delimited Format

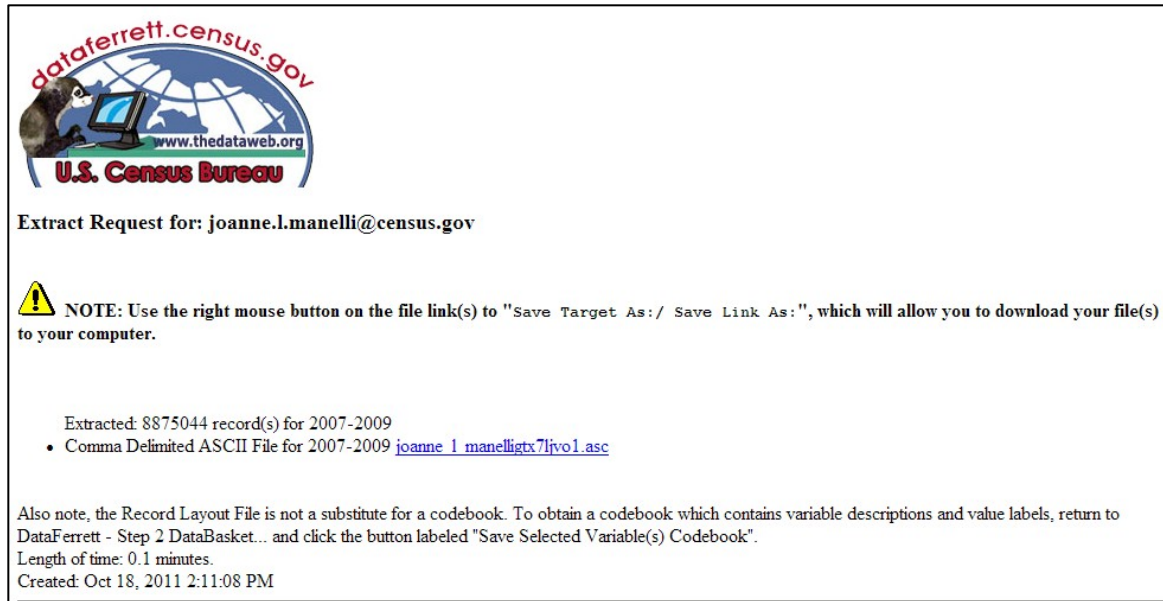
Space Delimited: A data file with each variable in a fixed starting position with a fixed length. Therefore, two files are available for download: the data file and a file (with file name ending in ".lay") containing the record layout information. The example below shows the results page for the information returned on the extract results page for this format type:



The screenshot shows the DataFerrett interface for a Space Delimited extract request. At the top is the DataFerrett logo with the URL dataferrett.census.gov and the U.S. Census Bureau logo. Below the logo, it says "Extract Request for: joanne.l.manelli@census.gov". A yellow warning icon is followed by a note: "NOTE: Use the right mouse button on the file link(s) to 'Save Target As:/ Save Link As:', which will allow you to download your file(s) to your computer." Below this is a bullet point: "Extracted: 8875044 record(s) for 2007-2009". Below that are two bullet points: "Space Delimited ASCII File for 2007-2009 [joanne l manelligtx79nfz1.asc](#)" and "Record Layout File for 2007-2009 [joanne l manelligtx79nfz1.lay](#)". At the bottom, it says: "Also note, the Record Layout File is not a substitute for a codebook. To obtain a codebook which contains variable descriptions and value labels, return to DataFerrett - Step 2 DataBasket... and click the button labeled 'Save Selected Variable(s) Codebook'." The length of time is 0.2 minutes and it was created on Oct 18, 2011 2:01:56 PM.

Figure 7-7: Extract Request for Space Delimited Format

Comma Delimited: The example below shows the information returned on the extract results page for this format type:



The screenshot shows the DataFerrett interface. At the top is a logo with a raccoon and the text "dataferrett.census.gov", "www.thedataweb.org", and "U.S. Census Bureau". Below the logo, it says "Extract Request for: joanne.l.manelli@census.gov". A yellow warning icon is followed by a note: "NOTE: Use the right mouse button on the file link(s) to 'Save Target As:/ Save Link As:', which will allow you to download your file(s) to your computer." Below this, it states "Extracted: 8875044 record(s) for 2007-2009" and lists a bullet point: "Comma Delimited ASCII File for 2007-2009 [joanne 1 manelligtx7ljvo1.asc](#)". At the bottom, there is additional text: "Also note, the Record Layout File is not a substitute for a codebook. To obtain a codebook which contains variable descriptions and value labels, return to DataFerrett - Step 2 DataBasket... and click the button labeled 'Save Selected Variable(s) Codebook'." followed by "Length of time: 0.1 minutes." and "Created: Oct 18, 2011 2:11:08 PM".

Figure 7-8: Extract Request for Comma Delimited Format

Specification Question 3: Compressed or Un-compressed?

Do you want to compress the data? This is only needed when the anticipated size of the file is quite large, possibly affecting the performance of the actual download to your computer. “No compression” is the default choice. Two commonly used types of compression are available if you choose to download the data in a compressed (aka “zip”) format: PKZIP (recognized by WinZip software) and the GZIP (aka “GNU Zip”) which runs on PC and Linux platforms.

The examples below show the types of file names you will see when using these two types of compression.



The screenshot shows the DataFerrett interface. At the top is a logo with a raccoon and the text "dataferrett.census.gov", "www.thedataweb.org", and "U.S. Census Bureau". Below the logo, it says "Extract Request for: joanne.l.manelli@census.gov". A yellow warning icon is followed by a note: "NOTE: Use the right mouse button on the file link(s) to 'Save Target As:/ Save Link As:', which will allow you to download your file(s) to your computer." Below this, it states "Extracted: 8875044 record(s) for 2007-2009" and lists a bullet point: "Zip Compressed Data File for 2007-2009 [joanne 1 manelligtx7ovil1.zip](#)". Below this, it lists another bullet point: "Record Layout File for 2007-2009 [joanne 1 manelligtx7ovil1.lay](#)".

Figure 7-9: Extract Request for PKZIP Format

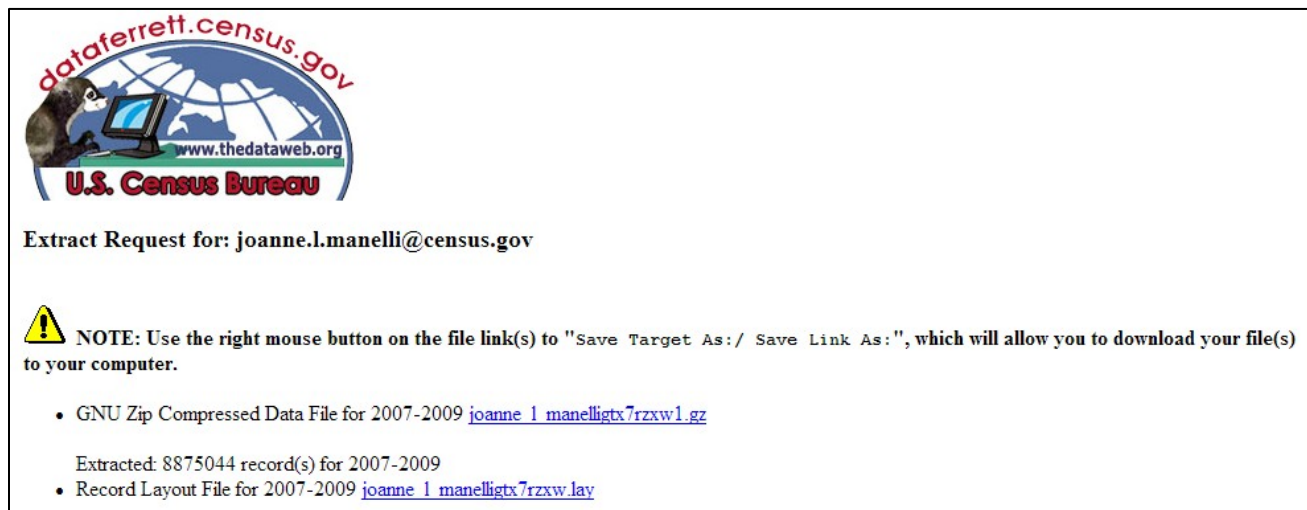


Figure 7-10: Extract Request for GZIP Format

Specification Question 4: Batch Mode?

The default is “No;” however, this option will interrupt processing your download request. If you want to be able to continue working while the download is processing, choose “Yes.” Choosing “Yes” will produce a pop-up message like the one shown in Figure 7-11 below. This message states that an email message will be sent to the email used when logging into the session of DataFerrett. This is one reason the correct input of your login email address is important. The last line in the message (truncated in Figure 7-11) contains the URL to use to download the data file to your computer. (**NOTE:** For the ACS 5-year Summary File, the extract request automatically runs in batch mode. You cannot change this.)

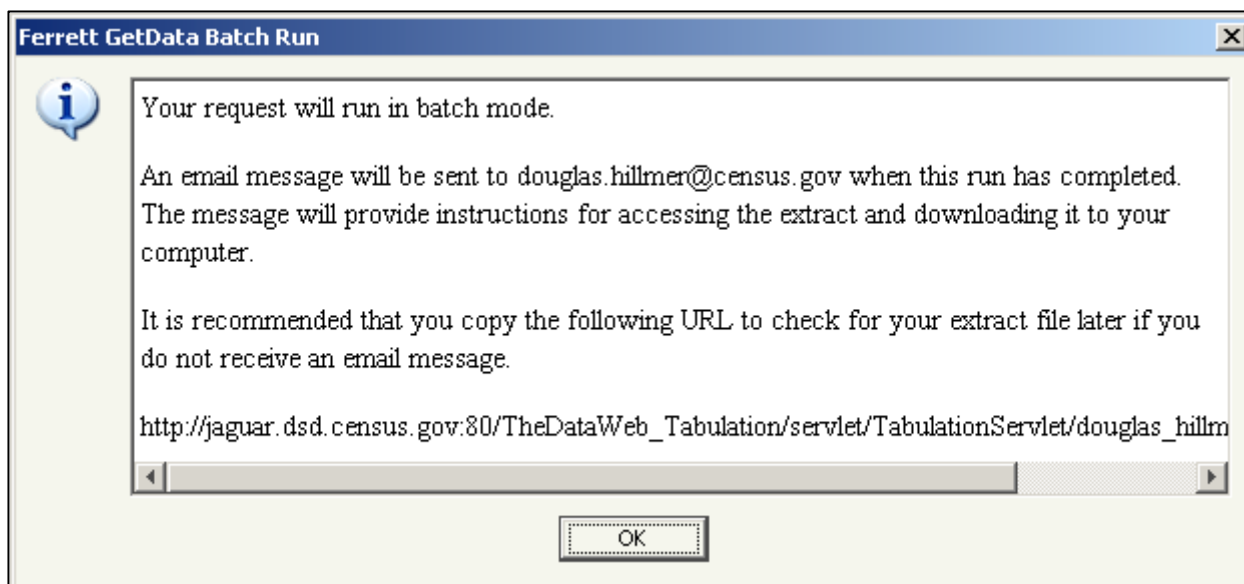


Figure 7-11: Message that Appears when “Batch” Method is Chosen for Download Processing

Subsequently, a message from “Support” will appear in your email inbox. The content of the message will be similar to:

“Your DataFerrett batch job completed.

To display your output filename(s) please click on the link below:

http://smpbff1.dsd.census.gov/TheDataWeb_Tabulation/servlet/TabulationServlet/douglas_hillmerq191rlf-scr.html

If you are unable to open the link directly, then copy and paste the URL above into your browser.”

Clicking on the link contained in the email message will use your default web browser to open a web page that will look similar to the one shown in Fig. 7-12.



Figure 7-12: Web Page with Link to Extract Request and Instructions for Downloading to Your Computer

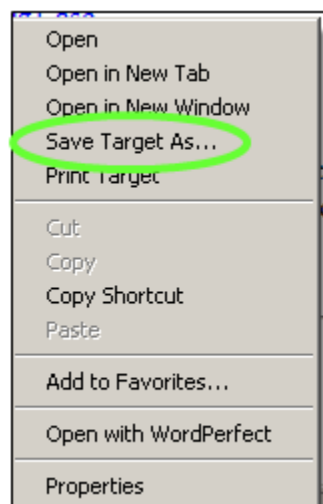


Figure 7-13: Menu Displaying “Save Target As” Selection

To save/download this data, use the right mouse button on the file link. Depending on the web browser, the menu that appears will contain a choice “Save Target As” or “Save Link As.” The image shown in Fig. 7-13 is displayed when using Internet Explorer as your web browser. Once you have right clicked on the file link and selected the action “Save Target as,” follow the instructions to download the file(s) to your computer.

WARNING: DO NOT save to favorites or bookmark. The files containing your extraction are deleted after ?? hours from the area referenced by the URL address. Therefore, it is safest to download the file(s) to your computer as soon as you get the email message from DataFerrett.

If you open the link (i.e., click on the left mouse button) it will show you the raw data in the number of rows selected. You can also see the data this way by selecting “Display data on the screen” instead of “Download Data” on the Download Data screen (see Fig. 7-3). Fig. 7-4 gives an example of such a display of the first records in an extract file.

Size Thresholds for Data Downloads in DataFerrett

DataFerrett limits the number of variables you can download and the size (in bytes) of the file. These limits also vary somewhat by the dataset being downloaded, as shown in the descriptions provided below.

Limits to the number of variables in a download:

ACS PUMS download limit Message: We’re sorry, the system only supports downloading 130 variables from this dataset. You have selected 282 variables. You will need to go back and reduce the number of variables.

CPS Basic PUMS selected all variables Message: We’re sorry, the system only supports downloading 50 variables from this dataset. You have selected 281 variables. You will need to go back and reduce the number of variables.

ACS 2005-2009 Summary File Message: We’re sorry, the system only supports downloading 528 variables from this dataset. You have selected 654 variables. You will need to go back and reduce the number of variables.

Other Limits:

There are three other types of file limitations. Figure 7-14 shows the pop-up window that appears when a user goes beyond any of these limits.

Limit to the number of geographic areas:

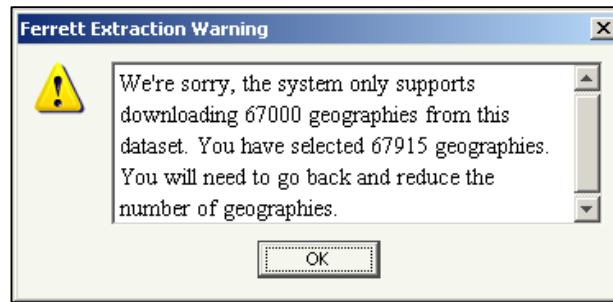


Figure 7-14: DataFerrett Extract Warning Limiting Number of Geographic Areas

Limit to the number of table cells in DataBasket:

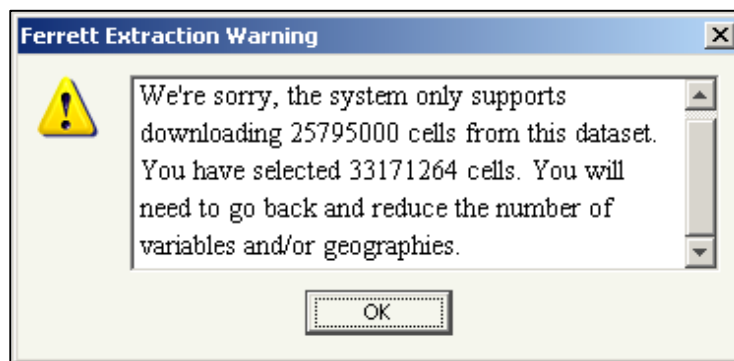


Figure 7-15: DataFerrett Extract Warning Limiting Number of Table Cells

If the file selected for the extract is very large, you will be prompted to select a file compression method from the methods available in the Download window:

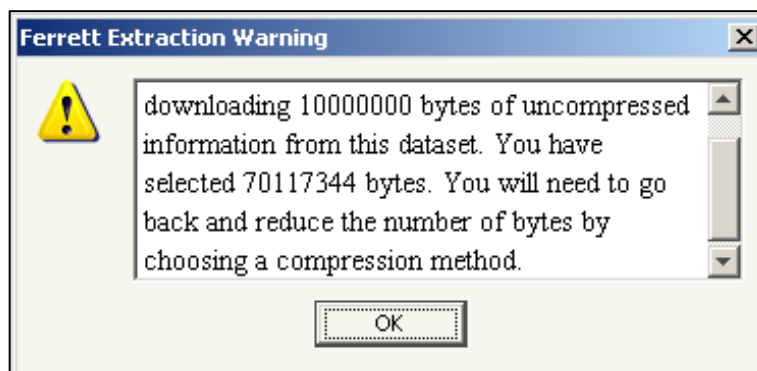


Figure 7-16: DataFerrett Extract Warning on Choosing Compression Method